(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 795 851 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
03.07.2002 Bulletin 2002/27

(51) Int Cl.[7]: **G10L 15/00**, G10L 15/28, G10L 17/00, G10L 11/00

(21) Application number: **97104406.0**

(22) Date of filing: **14.03.1997**

(54) Method and system for microphone array input type speech analysis

Verfahren und System zur Sprachanalyse mit Eingabe über eine Mikrophonanordnung

Procédé et système d'analyse de la parole à type d'entrée par réseau de microphones

(84) Designated Contracting States:
DE FR GB

(30) Priority: **15.03.1996 JP 5982196**

(43) Date of publication of application:
17.09.1997 Bulletin 1997/38

(73) Proprietor: **KABUSHIKI KAISHA TOSHIBA**
Kawasaki-shi, Kanagawa-ken 210-8572 (JP)

(72) Inventor: **Nagata, Yoshifumi**
Kobe-shi, Hyogo-ken (JP)

(74) Representative: **Zangs, Rainer E., Dipl.-Ing. et al**
Hoffmann Eitle,
Patent- und Rechtsanwälte,
Arabellastrasse 4
81925 München (DE)

(56) References cited:

- **SULLIVAN T M ET AL: "MULTI-MICROPHONE CORRELATION-BASED PROCESSING FOR ROBUST SPEECH RECOGNITION" SPEECH PROCESSING, MINNEAPOLIS, APR. 27 - 30, 1993, vol. VOL. 2, no. -, 27 April 1993, pages II-91-94, XP000427733 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS**
- **HUANG J ET AL: "A BIOMIMETIC SYSTEM FOR LOCALIZATION AND SEPARATION OF MULTIPLE SOUND SOURCES" IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, vol. 44, no. 3, 1 June 1995, pages 733-738, XP000527552**
- **BUB U ET AL: "KNOWING WHO TO LISTEN TO IN SPEECH RECOGNITION: VISUALLY GUIDED BEAMFORMING" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), DETROIT, MAY 9 - 12, 1995 SPEECH, vol. VOL. 1, 9 May 1995, pages 848-851, XP000670541 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art.

EP 0 795 851 B1

Description

[0001] The present invention relates to a microphone array input type speech recognition scheme in which speeches uttered by a user are inputted through microphone array and recognized.

[0002] In the speech recognition, a surrounding environment under which the speech input is made can largely affect the recognition performance. In particular, background noises and reflected sounds of the user's speeches can degrade the recognition performance so that they are sources of a serious problem encountered in a use of a speech recognition system. For this reason, in general, a short range microphone designed for use near the mouth of the user such as a headset microphone or a hand microphone has been employed, but it is uncomfortable to wear the headset microphone on a head for any extended period of time, while the hand microphone can limit a freedom of the user as it occupies the user's hands, and there has been a demand for a speech input scheme that can allow more freedom to the user.

[0003] A microphone array has been studied as a potential candidate for a speech input scheme that can resolve the conventionally encountered inconvenience described above, and there are some recent reports of its application to the speech recognition system. The microphone array is a set of a plurality of microphones which are arranged at spatially different positions, where noises can be reduced by the synthetic processing of outputs of these microphones.

[0004] Fig. 1 shows a configuration of a conventional speech recognition system using a microphone array. This speech recognition system of Fig. 1 comprises a speech input unit 11 having a microphone array formed by a plurality (N sets) of microphones, a sound source direction estimation unit 12, a sound source waveform estimation unit 13, a speech detection unit 14, a speech analysis unit 15, a pattern matching unit 16, and a recognition dictionary 17.

[0005] In this configuration of Fig. 1, the speech entered at the microphone array is converted into digital signals for respective microphones by the speech input unit 11, and the speech waveforms of all channels are entered into the sound source direction estimation unit 12.

[0006] At the sound source direction estimation unit 12, a sound source position or direction is estimated from time differences among signals from different microphones, using the known delay sum array method or a method based on the cross-correlation function as disclosed in U. Bub, et al.: "Knowing Who to Listen to in Speech Recognition: Visually Guided Beamforming", ICASSP '95, pp. 848-851, 1995.

[0007] A case of estimating a direction of the sound source and a case of estimating a position of the sound source respectively correspond to a case in which the sound source is far distanced from the microphone array so that the incident sound waves can be considered as plane waves and a case in which the sound source is relatively close to the microphone array so that the sound waves can be considered as propagating in forms of spherical waves.

[0008] Next, the sound source waveform estimation unit 13 focuses the microphone array to the sound source position or direction obtained by the sound source direction estimation unit 12 by using the delay sum array method, and estimates the speech waveform of the target sound source.

[0009] Thereafter, similarly as in the usual speech recognition system, the speech analysis is carried out for the obtained speech waveform by the speech analysis unit 15, and the pattern matching using the recognition dictionary 17 is carried out for the obtained analysis parameter, so as to obtain the recognition result. For a method of pattern matching, there are several known methods including the HMM (Hidden Markov Model), the multiple similarity method, and the DP matching, as detailed in Rabiner et al.: "Fundamentals of Speech Recognition", Prentice Hall, for example.

[0010] Now, in the speech recognition system, it is custom to input the speech waveform. For this reason, even in the conventional speech recognition system using the microphone array as described above, the sound source position (or the sound source direction) and the speech waveform are obtained by processing the microphone array outputs according to the delay sum array method, due to a need to estimate the speech waveform by a small amount of calculations. The delay sum array method is often utilized because the speech waveform can be obtained by a relatively small amount of calculations, but the delay sum array method is also associated with a problem that the separation power is lowered when a plurality of sound sources are located close to each other.

[0011] On the other hand, as a method for estimating the sound source position (or direction), there is a parametric method based on a model as disclosed in S. V. Pillai: "Array Signal Processing", Springer-Verlag, New York, 1989, for example, which is presumably capable of estimating the sound source position at higher precision than the delay sum array method, and which is also capable of obtaining the power spectrum necessary for the speech recognition from the sound source position estimation processing at the same time.

[0012] Fig. 2 shows a processing configuration for this conventionally proposed parametric method. In the configuration of Fig. 2, signals from a plurality of microphones are entered at a speech input unit 21, and the frequency analysis based on the FFT (Fast Fourier Transform) is carried out at a frequency analysis unit 22. Then, the sound source position estimation processing is carried out for each frequency component at a power estimation unit 23, and the final sound source position estimation result is obtained by synthesizing the estimation results for all the frequencies at a sound source direction judgment unit 24.

[0013] Here, the sound source position estimation processing is a processing for estimating a power at each direction or position while minutely changing a direction or position over a range in which the sound source can possibly be

located, so that a very large amount of calculations are required. In particular, in a case of assuming the propagation of sound waves in forms of spherical waves, it is going to estimate a position of the sound source rather than an arriving direction of the sound waves, so that two- or three-dimensional scanning is necessary and consequently an enormous amount of calculations are required.

[0014] Moreover, in the conventionally proposed parametric method described above, it is necessary to carry out this scanning processing for each frequency component obtained by the fast Fourier transform of the speech, so that it is difficult to reduce a required amount of calculations.

[0015] The document "Multi-microphone correlation based processing for robust recognition", by Thomas M. Sullivan and Richard M. Stern, IEEE Speech processing, Minneapolis, 1993, pages II - 91 to II - 94 discloses a method of signal processing for robust speech recognition using multiple microphones. Signals from each microphone are entered through a bank of band-pass filters, and the output of the band-pass filters is then entered into non linear rectifiers for rectifying the shape of the entered signal. The outputs from each rectifier are then cross-correlated within each frequency band.

[0016] It is therefore an object of the present invention to provide a method and a system for microphone array input type speech recognition capable of realizing a high precision sound source position or direction estimation by a small amount of calculations, and thereby realizing a high precision speech recognition. This object is achieved by obtaining a band-pass waveform, which is a waveform for each frequency bandwidth, from input signals of the microphone array, and directly obtaining a band-pass power of the sound source from the band-pass waveform. Then, the obtained band-pass power can be used as the speech parameter.

[0017] It is another object of the present invention to provide a method and a system for microphone array input type speech recognition capable of realizing the sound source estimation and the band-pass power estimation at high precision while further reducing an amount of calculations. This object is achieved by utilizing a sound source position search processing in which a low resolution position estimation and a high resolution position estimation are combined.

[0018] These objects are solved by the system of claim 1 and the method of claim 12. Advantageous embodiments are described in the dependent claims.

[0019] According to one aspect of the present application there is provided a microphone array input type speech recognition system, comprising: a speech input unit for inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; a frequency analysis unit for analyzing an input speech of each channel inputted by the speech input unit, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth; a sound source position search unit for calculating a band-pass power distribution for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the frequency analysis unit, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution; a speech parameter extraction unit for extracting a speech parameter for speech recognition, from the band-pass power distribution for each frequency bandwidth calculated by the sound source position search unit, according to the sound source position or direction estimated by the sound source position search unit; and a speech recognition unit for obtaining a speech recognition result by matching the speech parameter extracted by the speech parameter extraction unit with a recognition dictionary.

[0020] According to another aspect of the present application there is provided a microphone array input type speech analysis system, comprising: a speech input unit for inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; a frequency analysis unit for analyzing an input speech of each channel inputted by the speech input unit, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth; a sound source position search unit for calculating a band-pass power distribution for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the frequency analysis unit, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution; and a speech parameter extraction unit for extracting a speech parameter from the band-pass power distribution for each frequency bandwidth estimated by the sound source position search unit, according to the sound source position or direction estimated by the sound source position search unit.

[0021] According to another aspect of the present application there is provided a microphone array input type speech analysis system, comprising: a speech input unit for inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; a frequency analysis unit for analyzing an input speech of each channel inputted by the speech input unit, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth; and a sound source position search unit for calculating a band-pass power distribution for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the frequency analysis unit, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution.

[0022] According to another aspect of the present application there is provided a microphone array input type speech

recognition method, comprising the steps of: inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; analyzing an input speech of each channel inputted by the inputting step, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth; calculating a band-pass power distribution for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the analyzing step, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution; extracting a speech parameter for speech recognition, from the band-pass power distribution for each frequency bandwidth calculated by the calculating step, according to the sound source position or direction estimated by the calculating step; and obtaining a speech recognition result by matching the speech parameter extracted by the extracting step with a recognition dictionary.

[0023] According to another aspect of the present application there is provided a microphone array input type speech analysis method, comprising the steps of: inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; analyzing an input speech of each channel inputted by the inputting step, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth; calculating a band-pass power distribution for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the analyzing step, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution; and extracting a speech parameter from the band-pass power distribution for each frequency bandwidth calculated by the calculating step, according to the sound source position or direction estimated by the calculating step.

[0024] According to another aspect of the present application there is provided a microphone array input type speech analysis method, comprising the steps of: inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; analyzing an input speech of each channel inputted by the inputting step, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth; and calculating a band-pass power distribution for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the analyzing step, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution.

[0025] Other features and advantages of the present invention will become apparent from the following description taken in conjunction with the accompanying drawings.

[0026] Fig. 1 is a block diagram of a conventional microphone array input type speech recognition system.

[0027] Fig. 2 is a block diagram of a processing configuration for a conventionally proposed parametric method for estimating the sound source position or direction.

[0028] Fig. 3 is a block diagram of a microphone array input type speech recognition system according to the first embodiment of the present invention.

[0029] Fig. 4 is a diagram of a filter function to be used in a sound source position search unit in the system of Fig. 3.

[0030] Figs. 5A and 5B are diagrams showing a relationship between a sound source position and microphone positions in the system of Fig. 3, for a case of direction estimation and for a case of position estimation, respectively.

[0031] Fig. 6 is a diagram for explaining a peak detection from a sound source power distribution after the synthesizing processing in the system of Fig. 3.

[0032] Fig. 7 is a block diagram of one exemplary configuration for a speech recognition unit in the system of Fig. 3.

[0033] Fig. 8 is a block diagram of another exemplary configuration for a speech recognition unit in the system of Fig. 3.

[0034] Fig. 9 is a flow chart for the overall processing in the system of Fig. 3.

[0035] Fig. 10 is a diagram for explaining a read out waveform data for one frame used in a calculation of a correlation matrix at a sound source position search unit in the system of Fig. 3.

[0036] Figs. 11A and 11B are diagrams showing a relationship between a resolution and an increment value for search in a sound source position estimation in the system of Fig. 3.

[0037] Fig. 12 is a block diagram of a sound source position search unit in the system of Fig. 3 according to the second embodiment of the present invention.

[0038] Fig. 13 is a flow chart for the processing of the sound source position search unit of Fig. 12.

[0039] Referring now to Fig. 3 to Fig. 10, the first embodiment of a method and a system for microphone array input type speech recognition according to the present invention will be described in detail.

[0040] Fig. 3 shows a basic configuration of a microphone array input type speech recognition system in this first embodiment. This speech recognition system of Fig. 1 comprises a speech input unit 1, a frequency analysis unit 2, a sound source position search unit 3, a speech parameter extraction unit 4, and a speech recognition unit 5.

[0041] The speech input unit 1 has a microphone array (not shown) formed by N sets (8 sets, for example) of microphones, and converts speeches entered from the microphone array into digital signals.

[0042] The frequency analysis unit 2 analyzes the input speech for each microphone (channel) entered at the speech input unit 1 by a band-pass filter bank (a group of band-pass filters), and obtains a band-pass waveform which is a waveform for each frequency bandwidth.

[0043] The sound source position search unit 3 estimates a power arriving from each position or direction for each bandwidth according to the band-pass waveform for each frequency bandwidth obtained for each channel by the frequency analysis unit 2, as a sound source position judgement information, and identifies a sound source direction or position by synthesizing the obtained sound source position judgement information for a plurality of frequency bandwidths.

[0044] The speech parameter extraction unit 4 extracts the band-pass power of the speech signals arrived from the sound source direction or position identified by the sound source position search unit 3, as a speech parameter, according to the sound source position judgement information obtained at the sound source position search unit 3.

[0045] The speech recognition unit 5 carries out the speech recognition by matching the speech parameter extracted by the speech parameter extraction unit 4 with a recognition dictionary.

[0046] Now, the outline of the overall operation in the speech recognition system of Fig. 3 will be described.

[0047] First, the speeches entered at N (=8) sets of microphones are AD converted at the sampling frequency such as 12 KHz for example, microphone channel by microphone channel, at the speech input unit 1. Then, the frequency analysis is carried out at the frequency analysis unit 2, to obtain band-pass waveforms for a plurality of bands (bandwidths) corresponding to microphones. Here, a number M of bands used in the analysis is assumed to be equal to 16. The transmission bandwidths of the band-pass filters are to be determined as those required at the speech recognition unit 5. A manner of constructing the band-pass filters is well known and can be found in Rabiner et al.: "Fundamentals of Speech Recognition", Prentice Hall, for example.

[0048] Next, at the sound source position search unit 3, an arriving power in each bandwidth is estimated for each position or direction, according to the band-pass waveforms for N (=8) channels in each bandwidth outputted by the frequency analysis unit 2, as the sound source position judgement information. This processing is repeated for M (=16) times. This calculation of the sound source position judgement information is a calculation of an arriving power while sequentially displacing an assumed sound source position or direction, so as to obtain a distribution of arriving powers over a range in which the sound source can be located.

[0049] Thereafter, the sound wave arriving direction or the sound source position is estimated by synthesizing the above described power distribution obtained for each of M frequency bandwidths. Here, a position or direction with a large value in the power distribution is to be estimated as that of the source source.

[0050] In addition, at the speech parameter extraction unit 4, the sound source power (band-pass power) at the sound wave arriving direction or the sound source position is extracted, from the sound source position judgement information estimated for each bandwidth at the sound source position search unit 3, as the speech parameter. This speech parameter is then given to the speech recognition unit 5, where the speech recognition result is obtained and outputted.

[0051] As described, in this first embodiment, the sound source position is determined according to the estimated power distribution for each frequency bandwidth to be used in the speech recognition, and the speech parameter is obtained according to the determined sound source position, so that even when the sound source position is unknown, it is possible to realize the speech recognition by directly obtaining the speech parameter at high precision with a small amount of calculations.

[0052] Note that, when the sound source position is known, it suffices to obtain the arriving power value by limiting a power distribution calculation range to a single known sound source position or direction, and the configuration of Fig. 3 is also applicable to this case without any change. This simplified operation is effective when it is possible to assume that a user makes a speech input by approaching to a specific location.

[0053] Next, the detailed operation for obtaining the power distribution from a plurality of band-pass waveforms at the sound source position search unit 3 will be described.

[0054] At the sound source position search unit 3, in order to obtain a power at each direction or position from a plurality (M sets) of band-pass waveforms, the calculation of the minimum variance method is carried out. The minimum variance method is well known and described in Haykin: "Adaptive Filter Theory", Prentice Hall, for example.

[0055] In this first embodiment, at a time of the sound source power estimation by the minimum variance method, in order to deal with signals having a certain bandwidth which is not a narrow bandwidth, a filter function as indicated in Fig. 4 is realized by calculation, where the band-pass waveforms of the same frequency bandwidth obtained for N sets of microphones (i = 1 to N) by the frequency analysis unit 2 are added by an adder 32 after passing through transversal filters 31-1 to 31-N with multiple delay line taps corresponding to N sets of microphones (i = 1 to N). Here, the filter coefficients w_{11} to w_{1J}, ·····, w_{N1} to w_{NJ} of the filters 31-1 to 31-N are switchably set up for each bandwidth, so as to realize the filter function for all bandwidths.

[0056] In the configuration of Fig. 4, a number of taps in the filter is denoted as J, and a filter coefficient of the i-th microphone (microphone No. i) is denoted as w_{ij} ($1 \leqq i \leqq N$, $1 \leqq j \leqq J$). Here, J is equal to 10, for example, but this

setting may be be changed depending on a width of the bandwidth.

[0057] The filter output y in this configuration of Fig. 4 can be expressed as follows.

[0058] First, by denoting the band-pass waveform of the k-th frequency bandwidth ($1 \leqq k \leqq M$) obtained from the waveform for the i-th microphone as $x_{ik}(n)$, and by arranging band-pass waveform sample sequences: $x_{ik} = (x_{ik}(n-J+1), x_{ik}(n-J+2), \cdots\cdots, x_{ik}(n-1), x_{ik}(n))$ from J samples past of a certain time n up to this time n, for all N sets of microphones, it is possible to obtain a vector given by the following equation (1).

$$x_k = (x_{1k}, x_{2k}, \cdots\cdots, x_{Nk})^T \tag{1}$$

[0059] Also, by arranging the filter coefficients w_{ij}, it is possible to obtain a vector given by the following equation (2).

$$w_k = (w_{11}, w_{12}, \cdots\cdots, w_{1J}, w_{21}, w_{22}, \cdots\cdots, w_{2J}, \cdots\cdots, w_{N1}, w_{N2}, \cdots\cdots, w_{NJ})^T \tag{2}$$

[0060] Using the above equations (1) and (2), the filter output y can be expressed as:

$$y = w_k \cdot x_k \tag{3}$$

where · denotes a complex conjugate of a vector. In this expression, x_k is usually called a snap-shot.

[0061] Now, denoting the expectation value as E[ ], the expectation value of the filter output power y^2 is expressed as :

$$E[y^2] = E[w_k \cdot x_k x_k \cdot w_k] = w_k \cdot R_k w_k \tag{4}$$

where $R_k = E[x_k x_k^*]$ is a correlation matrix of x. Then, the estimation vector according to the minimum variance method is obtained by minimizing this expectation value $E[y^2]$ under the constraint conditions that a response of the microphone array for a target direction or position is to be maintained constant.

[0062] These constraint conditions can be expressed as:

$$w_k \cdot A = g \tag{5}$$

where g is a column vector with constant values in a size equal to a number L of constraint conditions. For example, this g can be [1, 1, ·····, 1]. Also, A is a matrix formed by direction control vectors am for different frequencies as column vectors. This matrix A can be expressed as:

$$A = [a_1, a_2, \cdots\cdots, a_L] \tag{6}$$

and each direction control vector a_m (m = 1, 2, ·····, L) can be expressed as:

$$a_m = (1, a_2 e^{-j\omega_m \tau_2}, \cdots\cdots, a_N e^{-j\omega_m \tau_N}) \tag{7}$$

where $\tau 2, \cdots\cdots, \tau_N$ are propagation time differences of the incident sound wave for the second to N-th microphones with reference to the first microphone, respectively. Note that the propagation time difference τ_1 of the incident sound wave for the first microphone is set equal to zero. Also, ω_m is an angular frequency, and $a_2, \cdots\cdots, a_N$ are amplitude ratio of the incident sound wave for the second to N-th microphone with reference to the first microphone, respectively. Here, L Is set equal to 10, for example, and ω_m is set to be $\omega_m = ((\omega_a - \omega_b)/(L-1)) \cdot m + \omega_b$, where ω_a is an upper limit angular frequency of the bandwidth and ω_b is a lower limit angular frequency of the bandwidth.

[0063] When the problem of minimization under the constraints given by the equations (4) and (5) is solved by the Lagrange's method of indeterminate coefficients, the filter coefficient w_k for minimizing the arriving power from any direction or position other than the sound source direction θ or the sound source position θ can be given by the following

equation (8).

$$w_k = R_k^{-1} A (A^{\cdot} R_k^{-1} A)^{-1} g \quad (8)$$

[0064] Using this filter coefficient w_k, the arriving power (arriving band-pass power, sound source power) $P_k(\theta)$ for the k-th bandwidth from the sound source θ can be calculated as the following equation (9).

$$P_k(\theta) = g^{\cdot} (A^{\cdot} R_k^{-1} A)^{-1} g \quad (9)$$

In a case of the sound source position estimation, θ is taken as a vector for expressing the coordinates.

[0065] Now, with reference to Figs. 5A and 5B, a manner of obtaining the propagation time difference and the amplitude for each microphone will be described. Here, the explanation is given on two-dimensional plane for the sake of simplicity, but the extension to the three-dimensional space should be obvious.

[0066] First, as shown in Fig. 5A, the coordinates of the first microphone (No. 1) are denoted as (x_1, y_1) and the coordinates of the i-th microphone (No. i) are denoted as (x_i, y_i). Then, for a case of the plane waves, when the sound waves are incident from a direction θ, the propagation time difference τ_i of the incident sound waves at the i-th microphone and the first microphone is given by:

$$\tau_i(\theta) = ((x_i-x_1)^2 + (y_i-y_1)^2)^{1/2} \cos(\theta-\tan^{-1}((y_i-y_1)/(x_i-x_1))) \quad (10)$$

and the amplitude can be assumed as:

$$a_1 = a_2 = \cdots\cdots = a_N = 1 \quad (11)$$

[0067] On the other hand, in a case of a point sound source, as shown in Fig. 5B, when an assumed sound source position θ is located at (x_s, y_s), the propagation time difference τ_i and the amplitude a_i can be given by:

$$\tau_i = (((x_i- x_s)^2 + (y_i-y_s)^2)^{1/2} -((x_i-x_s)^2 + (y_i-y_s)^2)^{1/2})/c \quad (12)$$

and

$$a_i = ((x_i-x_s)^2 + (y_i-y_s)^2)^{1/2} /((x_i-x_s)^2 + (y_i-y_s)^2)^{1/2} \quad (13)$$

where c is the speed of sound.

[0068] $P_k(\theta)$ given by the above equation (9) becomes large when θ coincides with the arriving direction or the sound source position, or small when they do not coincide. For this reason, by calculating $P_k(\theta)$ for each direction or position, the arriving direction or the sound source position can be estimated as a position of the peak.

[0069] To this end, in a case of obtaining the sound source direction, the sound source position search unit 3 calculates $P_k(\theta)$ while sequentially changing θ gradually, 1° by 1° for example. Also, in a case of obtaining the sound source position, the sound source position search unit 3 calculates $P_k(\theta)$ for lattice points at 2 cm interval for example, within the search range. The increment value for θ may be changed to any appropriate value depending on factors such as a wavelength and a distance to an assumed sound source position.

[0070] Next, at the sound source position search unit 3, $P_k(\theta)$ of the equation (9) obtained for all bandwidths are synthesized to estimate the sound source position or the sound source direction.

[0071] Here, the synthesizing can be realized by multiplying a weight W_k to the arriving power distribution $P_k(\theta)$ for

each bandwidth, and taking a sum for all the frequency bandwidths from k = 1 to k = M, that is:

$$P(\theta)_{total} = \Sigma W_k P_k(\theta) \quad (14)$$

and estimating the sound source from a peak on the distribution after this synthesizing processing (the total sound source power distribution).

[0072] Here, all the weights W_k may be set equal to 1, or the weight for the frequency of a noise source with a known frequency characteristic such as a power source frequency may be set small so as to reduce the influence of the noise.

[0073] The detection of the sound source is carried out according to a size of a peak in $P(\theta)_{total}$ as described above, and a single largest peak can be detected as the sound source. Alternatively, as shown in Fig. 6, by setting a prescribed threshold with reference to an average value of portions other than peak portions on the synthesized (total) sound source power distribution, such as 5 dB, and all peaks above this threshold may be detected as the sound sources, while not detecting any sound source at all when there is no peak above this threshold.

[0074] In this manner, the arriving power distribution $P_k(\theta)$ for each bandwidth given by the equation (9) is used in Judging whether the sound source exists at an assumed sound source position which is set to be each direction or position determined with reference to positions of a plurality of microphones, so that this arriving power distribution $P_k(\theta)$ will be referred to as the sound source position judgement information.

[0075] Next, the speech parameter extraction unit 4 can extract the power of the k-th frequency bandwidth of the sound source, from the already obtained arriving power distribution $P_k(\theta)$ for each bandwidth, according to the sound source direction or position obtained by the sound source position search unit 3. Consequently, by extracting the power for all the bandwidths from k = 1 to k = M, it is possible to obtain the band-pass power to be used as the speech parameter.

[0076] The band-pass power of the sound source obtained in this manner is sent from the speech parameter extraction unit 4 to the speech recognition unit 5, and used in the speech recognition processing.

[0077] As shown in Fig. 7, the speech recognition unit 5 comprises a speech power calculation unit 501, a speech detection unit 502, a pattern matching unit 503, and a recognition dictionary 504.

[0078] In this speech recognition unit 5, the speech power is calculated from the speech parameter (band-pass power) extracted by the speech parameter extraction unit 4, and a speech section is detection by the speech detection unit 502 according to the calculated speech power. Then, for the speech parameter in the detected speech section, the pattern matching with the recognition dictionary 504 is carried out by the pattern matching unit 503.

[0079] Note that, as shown in Fig. 8, the speech recognition unit 5 may be formed by a pattern matching unit 511, a recognition dictionary 512, and a speech detection unit 513, so as to carry out the word spotting scheme in which the continuous pattern matching for the speech parameter is carried out and a section with the largest matching score is determined as a speech section.

[0080] The total speech power can be obtained by summing powers for all the bandwidths extracted by the speech parameter extraction unit 4, so that it is possible to use the known speech section detection method based on the speech power as disclosed in L. F. Lamel et al.: "An Improved Endpoint Detector for Isolated Word Recognition", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.ASSP-29, No. 4, pp. 777-785, August 1981. The above processing is to be carried out for each frame of the input speech waveform data so as to recognize the speech continuously.

[0081] In this first embodiment, using the configuration of the microphone array input type speech recognition system as described above, the speech recognition is realized by directly obtaining the band-pass power which is the speech parameter. Consequently, it is possible to use the minimum variance method which is a sound source direction or sound source position estimation method with a high precision and a small amount of calculations.

[0082] Now, the flow of processing as described above will be summarized with reference to the flow chart of Fig. 9.

[0083] First, prior to the start of the processing, the initial setting is made for factors such as whether a direction estimation is to be used or a position estimation is to be used, a range of sound source search to be used, and an increment value to be used in the search (step S1). In an example of Fig. 9, a direction estimation is to be used, a searching range is from -90° to +90°, and an increment value for search is 1°.

[0084] Next, the speeches entered at N sets of microphones are A/D converted at the sampling frequency of 12 KHz for example, for all N channels in parallel, by the speech input unit 1. The obtained waveform data are then stored in a buffer (not shown) of the speech input unit 1 (step S2). Normally, this step S2 is carried out continuously in real time, regardless of the other processing.

[0085] Next, the waveform data for each channel is read out from the buffer for one frame size, 256 points for example, and applied to the band-pass filter bank at the frequency analysis unit 2, so as to extract the band-pass waveform for each frequency bandwidth k (k = 1 to M), where M = 16 in this example (step S3). Here, the calculations for the band-

pass filters may be carried out independently for each microphone in parallel, or sequentially for each microphone serially.

[0086] Next, at the source source position search unit 3, using the band-pass waveform data for N channels obtained by the frequency analysis unit 2 at the step S3, the correlation matrix R_k for each frequency bandwidth k is obtained (step S4). Here, as shown in Fig. 10, the calculation of the correlation matrix R_k is realized by obtaining the correlation matrix R_k as a time average of the auto-correlation matrices of samples for 20 frames taken at an interval of 2 samples from a frame data of 256 samples (points) for example.

[0087] In addition, at the step S4, using this correlation matrix R_k, the arriving power distribution $Pk(\theta) = g \cdot (A \cdot R_k^{-1} A)^{-1} g$ is obtained as the sound source position judgement information for each assumed position or direction. This calculation is carried out over the entire space to be searched through, so as to obtain the spatial distribution of the arriving powers. As for the bandwidths, when M = 16, the calculation is carried out from k = 1 to k = 16.

[0088] Next, at the sound source position search unit 3, the arriving power distributions $P_k(\theta)$ for different frequency bandwidths are summed over the entire frequency bandwidths, for each θ, so as to obtain the total sound source power distribution $P(\theta)_{total}$. Then, the largest peak is extracted from this $P(\theta)_{total}$ and identified as the sound source position θ_0 (step S5).

[0089] Next, at the speech parameter extraction unit 4, a value on the arriving power distribution (sound source position judgement information distribution) $P_k(\theta)$ for each frequency bandwidth obtained by the sound source position search unit 3 at the sound source position θ_0 is extracted, and this is repeated for all the frequency bandwidths for each sound source, so as to obtain the speech parameter $P_k(\theta_0)$ (step S6).

[0090] In addition, at the step S6, the powers for different bandwidths k of the speech parameter $P_k(\theta)$ are summed to obtain the power for the entire speech bandwidth at the speech power calculation unit 501 of the speech recognition unit 5.

[0091] Next, using the power for the entire speech bandwidth obtained at the step S6, the speech section is detected by the speech detection unit 502 of the speech recognition unit 5 (step S7).

[0092] Then, whether the end of the speech section is detected by the speech detection unit 502 or not is judged (step S8), and if not, the processing returns to the step S2 to carry out the frequency analysis for the next waveform data frame.

[0093] On the other hand, when the end of the speech section is detected, a matching of the speech parameter in the detected speech section with the recognition dictionary 504 is carried out, and the obtained recognition result is outputted (step S9). Then, the processing returns to the step S2 to carry out the frequency analysis for the next waveform data frame.

[0094] Thereafter, the above processing is repeated so as to carry out the speech parameter estimation and the speech recognition continuously.

[0095] Note that the processing described above can be carried out at high speed by adopting the pipeline processing using a plurality of processors (as many processors as a number of microphones, for example) which are operated in parallel.

[0096] Referring now to Fig. 11A to Fig. 13, the second embodiment of a method and a system for microphone array input type speech recognition according to the present invention will be described in detail.

[0097] This second embodiment is directed to a scheme for further reducing an amount of calculations in the sound source position estimation, by changing the constraint conditions used in the spectrum estimation so as to control the resolution, and changing the search density according to the resolution so as to reduce an amount of calculation in the sound source search. In this second embodiment, the basic configuration of the speech recognition system is the same as that of the first embodiment, so that Fig. 3 will be also referred in the following description.

[0098] In the first embodiment described above, the constraint conditions used in the spectrum estimation based on the minimum variance method are that a response of the microphone array for one direction or position is to be maintained constant. In this case, the resolution of the estimation is sufficiently high, so that the peak is found by the dense search in which the arriving power is obtained while changing θ 1° by 1° for example within the sound source search range.

[0099] When the resolution is so high as in this case, as shown in Fig. 11A, there is a possibility for failing to detect an accurate apex of the peak when the search is carried out not so densely, so that an amount of calculations for the search cannot be reduced easily.

[0100] In contrast, when the resolution of the sound source position estimation processing can be lowered, as shown in Fig. 11B, a possibility for overlooking the peak position can be made low even when the search is carried out coarsely, so that an amount of calculations can be reduced. In this case, however, it may not be possible to separate the closely located sound sources, or the estimation precision may be lowered, as much as the resolution is lowered.

[0101] For this reason, this second embodiment adopts a scheme in which the lower resolution search is carried out first and then the high resolution search is carried out only in a vicinity of the peak, so as to realize the high precision sound source position estimation with a small amount of calculations. This scheme will now be described in detail.

[0102] The resolution at a time of the sound source position estimation can be controlled by requiring that responses of the microphone array for a plurality of directions or positions are to be simultaneously maintained constant, instead of just requiring a response of the microphone array for one direction or position to be maintained constant as in the constraint conditions of the equation (5).

[0103] For example, using two time delays $\tau_1(\theta_1)$ and $\tau_2(\theta_2)$ for two angles θ_1 and θ_2 (according to the equation (10)), it is possible to use two direction control vectors $a_m(\theta_1)$ and $a_m(\theta_2)$ (m = 1, 2, ·····, L) given by the following equations (15) and (16).

$$a_m(\theta_1) = (1, a_2 e^{-j\omega_m \tau_2(\theta_1)}, \cdots\cdots, a_N e^{-j\omega_m \tau_n(\theta_1)}) \quad (15)$$

$$a_m(\theta_2) = (1, a_2 e^{-j\omega_m \tau_2(\theta_2)}, \cdots\cdots, a_N e^{-j\omega_m \tau_N(\theta_2)}) \quad (16)$$

[0104] Then, using these two direction control vectors $a_m(\theta_1)$ and $a_m(\theta_2)$, it is possible to set:

$$A = [a_1(\theta_1), a_2(\theta_1), \cdots\cdots, a_L(\theta_1),$$
$$a_1(\theta_2), a_2(\theta_2), \cdots\cdots, a_L(\theta_2)] \quad (17)$$

so as to make responses of the microphone array to two directions simultaneously.

[0105] Here, when θ_1 and θ_2 are set to close values, such as $\theta_2 = \theta_1 + 1°$ for example, it is equivalent to a case of making a response of the microphone array for a single direction with a width between θ_1 and θ_2, so that it is equivalent to the lowering of the resolution. Note here that a number of directions for which the responses of the microphone array are to be made simultaneously is not necessarily limited to two.

[0106] When the resolution is lowered, the search can be coarser compared with a case of high resolution, so that an amount of calculations can be reduced.

[0107] Then, after the search using the low resolution sound source position estimation processing described above, the high resolution search as described in the first embodiment can be carried out only in a vicinity of the peak detected by the first search, so that the high precision sound source position estimation can be realized with a reduced amount of calculations overall.

[0108] Fig. 12 shows a configuration of the sound source position search unit 3 in this second embodiment for realizing the above described sound source position estimation processing.

[0109] In this configuration of Fig. 12, the sound source position search unit 3 comprises a low resolution sound source position search unit 301 and a high resolution sound source position search unit 302. The low resolution sound source position search unit 301 coarsely estimates the arriving power distribution in terms of positions or directions by using the low resolution spectrum estimation. The high resolution sound source position search unit 302 densely estimates the arriving power distribution by using the high resolution spectrum estimation only in a vicinity of the position or direction obtained by the low resolution sound source position search unit 301.

[0110] Now, the flow of processing in this sound source position search unit 3 in the configuration of Fig. 12 will be described with reference to the flow chart of Fig. 13.

[0111] First, using the inputs of the band-pass waveforms corresponding to the microphone, the correlation matrix R_k is calculated (step S11). Here, a method for obtaining this correlation matrix R_k is the same as in the first embodiment.

[0112] Next, using the obtained correlation matrix R_k, the low resolution sound source position search is carried out (step S12). At this point, the increment value θ_d for search is set to be a relatively large value, such as 5° for example, so as to carry out the search coarsely over the entire search range. Also, in order to lower the resolution, a matrix as expressed by the equation (17) which has two direction control vectors $a_m(\theta_1)$ and $a_m(\theta_2)$ for two directions or positions as expressed by the equations (15) and (16) as column vectors is used instead of a matrix A in the equation (9). In Fig. 13, this matrix is denoted as B in order to distinguish it from a matrix A of the equation (6). The search is carried out for each bandwidth.

[0113] Next, the low resolution arriving power distributions for different bandwidths are synthesized, and the sound source position θ_0 is obtained from a peak therein (step S13).

[0114] Next, in a vicinity of the sound source position obtained at the step S13, the high resolution sound source position search is carried out. Here, the setting of the search range is set to be ±10° of the sound source position obtained at the step S13, for example. At this point, the equation to be used for the arriving power estimation (arriving power distribution) is the same as the equation (9), and the increment value is set to a smaller value such as 1° for

example (step S14).

[0115] Next, the high resolution arriving power distributions for different bandwidths obtained at the step S14 are synthesized, and the sound source position θ_0 ' is obtained from a peak therein (step S15).

[0116] At the speech parameter extraction unit 4, the power (speech parameter) of the sound source is extracted from the arriving power distribution obtained by the high resolution sound source position search at the high resolution sound source position search unit 302 of the sound source position search unit 3.

[0117] As described, in this second embodiment, by using the sound source position search processing in which the low resolution sound source position estimation and the high resolution sound source position estimation are combined, it is possible to estimate the sound source position and its band-pass power while reducing an amount of calculations considerably.

[0118] As described, according to the present invention, a band-pass waveform which is a waveform for each frequency bandwidth is obtained from input signals of the microphone array, and a band-pass power of the sound source is directly obtained from the band-pass waveform, so that it is possible to realize a high precision sound source position or direction estimation by a small amount of calculations. Moreover, the obtained band-pass power can be used as the speech parameter so that it is possible to realize a high precision speech recognition.

[0119] In the speech recognition system of the present invention, the input signals of the microphone array entered by the speech input unit are frequency analyzed by the frequency analysis unit, so as to obtain the band-pass waveform which is a waveform for each frequency bandwidth. This band-pass waveform is obtained by using the band-pass filter band (a group of band-pass filters), instead of using the frequency analysis based on FFT as in the conventional speech recognition system. Then, the band-pass power of the sound source is directly obtained from the obtained band-pass waveform by the sound source position search unit.

[0120] Here, in order to handle signals within some bandwidth collectively, a filter configuration (filter function) having a plurality of delay line taps for each microphone channel is used and the sound source power is obtained as a sum of the filter outputs for all channels, while using the minimum variance method which is a known high precision spectrum estimation method.

[0121] The sound source power estimation processing using the minimum variance method is also used in the conventionally proposed parametric method described above, but a use of only one delay line tap has been assumed conventionally, so that it has been impossible to obtain the bandwidth power collectively.

[0122] In contrast, in the speech recognition system of the present invention, a filter configuration with a plurality of delay line taps is used so that the power in each direction or position is obtained for each frequency bandwidth necessary for the speech recognition, rather than obtaining the power in each direction or position for each frequency, and therefore the obtained power can be directly used for the speech recognition while a required amount of calculations can be reduced.

[0123] For example, in a case of using the conventional FFT with 512 points, it has been necessary to repeatedly obtain the power in each direction for each of 256 components, but in the present invention, when a number of bands in the band-pass filter bank is set to 16 for example, it suffices to estimate the power in each direction for 16 times. In addition, this power (band-pass power) can be estimated at higher precision compared with the conventional case of using the delay sum array processing, so that it is possible to realize the high precision speech recognition.

[0124] In the sound source position search unit, the synthesizing of the band-pass power distributions for a plurality of frequency bandwidths can be realized by multiplying respective weights to the band-pass power distributions of different bandwidths, and taking a sum of the weighted band-pass power distributions for all the frequency bandwidths. Here, all the weights may be set equal to 1, or the weight for the frequency of a noise source with a known frequency characteristic such as a power source frequency may be set small so as to reduce the influence of the noise.

[0125] Also, in the sound source position search unit, the estimation of the sound source position or direction from the synthesized power distribution for each position or direction can be realized by detecting a peak in the synthesized power distribution and setting a position or direction corresponding to the detected peak as the sound osurce position or direction.

[0126] Furthermore, according to the present invention, by using a sound source position search processing in which a low resolution position estimation and a high resolution position estimation are combined, it is possible to realize the sound source estimation and the band-pass power estimation at high precision while further reducing an amount of calculations.

[0127] In this case, the coarse search carried out by the sound source position search unit for the purpose of reducing an amount of calculations is the low resolution search so that it is possible to make it highly unlikely to overlook the sound source position or direction (a peak position in the power distribution).

[0128] Although there is a possibility for being unable to separate closely located sound sources or lowering the estimation precision by the low resolution search alone, the high resolution search is also carried out only in a vicinity of the sound source position or direction obtained by the low resolution search (in a vicinity of the detected peak), so that it is possible to realize the high precision sound source position estimation with a further reduced amount of cal-

culations.

[0129] It is to be noted that, in the above description, the speech recognition system having a speech recognition unit has been described, but the speech parameter extraction technique according to the present invention can be utilized separately from the speech recognition device.

[0130] Namely, it is also possible to provide a microphone array input type speech analysis system for analyzing the input signals of the microphone array and extracting the speech parameter, which is formed by the speech input unit, the frequency analysis unit, the sound source position search unit, and the speech parameter extraction unit substantially as described above, in which the band-pass waveform which is a waveform for each frequency bandwidth is obtained from input signals of the microphone array, a band-pass power of the sound source is directly obtained from the band-pass waveform, and the obtained band-pass power is used as the speech parameter.

[0131] Similarly, it is also possible to provide a microphone array input type speech analysis system for analyzing the input signals of the microphone array and estimating the sound source position or direction, which is formed by the speech input unit, the frequency analysis unit, and the sound source position search unit substantially as described above, in which the band-pass waveform which is a waveform for each frequency bandwidth is obtained from input signals of the microphone array, a band-pass power of the sound source is directly obtained from the band-pass waveform, and the sound source position or direction is estimated according to a synthesized band-pass power distribution.

[0132] Such a microphone array input type speech analysis system according to the present invention is utilizable not only in the speech recognition but also in the other speech related processing such as a speaker recognition.

[0133] It is also to be noted that, besides those already mentioned above, many modifications and variations of the above embodiments may be made without departing from the novel and advantageous features of the present invention. Accordingly, all such modifications and variations are intended to be included within the scope of the appended claims.

Claims

1. A microphone array input type speech analysis system, comprising:

   a speech input unit (1) for inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; and

   a frequency analysis unit (2) for analysing an input speech of each channel inputted by the speech input unit, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth;

   characterized by

   a sound source position search unit (3) for calculating a band-pass power distribution as a function of sound source position or direction for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the frequency analysis unit, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution.

2. The system of claim 1, further comprising:

   a speech parameter extraction unit (4) for extracting a speech parameter from the band-pass power distribution for each frequency bandwidth calculated by the sound source position search unit, according to the sound source position or direction estimated by the sound source position search unit.

3. The system of claim 2, further comprising:

   a speech recognition unit (5) for obtaining a speech recognition result by matching the speech parameter extracted by the speech parameter extraction unit with a recognition dictionary.

4. The system of any one of claims 1, 2 and 3, wherein the sound source position search unit includes:

   a low resolution sound source position estimation unit for estimating a rough sound source position or direction, by minimizing an output power of the microphone array under constraints that responses of the microphone array for a plurality of directions or positions are to be maintained constant; and

a high resolution sound source position estimation unit for estimating an accurate sound source position or direction in a vicinity of the rough sound source position or direction estimated by the low resolution sound source position estimation unit, by minimizing the output power of the microphone array under constraints that a response of the microphone array for one direction or position is to be maintained constant, wherein the speech parameter extraction unit extracts the speech parameter for speech recognition according to the accurate sound source position or direction.

5. The system of claim 3, wherein the frequency analysis unit obtains the band-pass waveforms for each channel by using a band-pass filter bank.

6. The system of claim 3, wherein the sound source position search unit calculates the band-pass power distribution for each frequency bandwidth, by calculating a band-pass power for each frequency bandwidth, in each one of a plurality of assumed sound source positions or directions within a prescribed search range.

7. The system of claim 3, wherein the sound source position search unit calculates the band-pass power distribution for each frequency bandwidth by using a filter function configuration having a plurality of delay line taps for each channel.

8. The system of claim 3, wherein the sound source position search unit calculates the band-pass power distribution for each frequency bandwidth by using a minimum variance method for minimizing an output power of the microphone array under constraints that a response of the microphone array for one direction or position is to be maintained constant.

9. The system of claim 3, wherein the speech parameter extraction unit extracts the band-pass power distribution for each frequency bandwidth calculated by the sound source position search unit for the sound source position or direction estimated by the sound source position search unit directly as the speech parameter.

10. The system of claim 3, wherein the sound source position search unit synthesizes the calculated band-pass power distributions for a plurality of frequency bandwidths by weighting the calculated band-pass power distributions with respective weights, and summing weighted band-pass power distributions.

11. The system of claim 3, wherein the sound source position search unit estimates the sound source position or direction by detecting a peak in the synthesized band-pass power distribution and setting a position or direction corresponding to a detected peak as the sound source position or direction.

12. A microphone array input type speech analysis method, comprising the steps of:

inputting speeches in a plurality of channels using a microphone array formed by a plurality of microphones; and

analysing an input speech of each channel inputted by the inputting step, and obtaining band-pass waveforms for each channel, each band-pass waveform being a waveform for each frequency bandwidth;

characterized by

calculating a band-pass power distribution as a function of sound source position or direction for each frequency bandwidth from the band-pass waveforms for each frequency bandwidth obtained by the analysing step, synthesizing calculated band-pass power distributions for a plurality of frequency bandwidths, and estimating a sound source position or direction from a synthesized band-pass power distribution.

13. The method of claim 12, further comprising the step of:

extracting a speech parameter from the band-pass power distribution for each frequency bandwidth calculated by the calculating step, according to the sound source position or direction estimated by the calculating step.

14. The method of claim 13, further comprising the step of:

obtaining a speech recognition result by matching the speech parameter extracted by the extracting step with a recognition dictionary.

15. The method of any one of claims 12, 13 and 14, wherein the calculating step includes the steps of:

a low resolution sound source position estimation step for estimating a rough sound source position or direction, by minimizing an output power of the microphone array under constraints that responses of the microphone array for a plurality of directions or positions are to be maintained constant; and

a high resolution sound source position estimation step for estimating an accurate sound source position or direction in a vicinity of the rough sound source position or direction estimated by the low resolution sound source position estimation step, by minimizing the output power of the microphone array under constraints that a response of the microphone array for one direction or position is to be maintained constant, wherein the extracting step extracts the speech parameter for speech recognition according to the accurate sound source position or direction.

16. The method of claim 14, wherein the analysing step obtains the band-pass waveforms for each channel by using a band-pass filter bank.

17. The method of claim 14, wherein the calculating step calculates the band-pass power distribution for each frequency bandwidth, by calculating a band-pass power for each frequency bandwidth, in each one of a plurality of assumed sound source positions or directions within a prescribed search range.

18. The method of claim 14, wherein the calculating step calculates the band-pass power distribution for each frequency bandwidth by using a filter function configuration having a plurality of delay line taps for each channel.

19. The method of claim 14, wherein the calculating step calculates the band-pass power distribution for each frequency bandwidth by using a minimum variance method for minimizing an output power of the microphone array under constraints that a response of the microphone array for one direction or position is to be maintained constant.

20. The method of claim 14, wherein the extracting step extracts the band-pass power distribution for each frequency bandwidth calculated by the calculating step for the sound source position or direction estimated by the calculating step directly as the speech parameter.

21. The method of claim 14, wherein the calculating step synthesizes the calculated band-pass power distributions for a plurality of frequency bandwidths by weighting the calculated band-pass power distributions with respective weights, and summing weighted band-pass power distributions.

22. The method of claim 14, wherein the calculating step estimates the sound source position or direction by detecting a peak in the synthesized band-pass power distribution and setting a position or direction corresponding to a detected peak as the sound source position or direction.

Patentansprüche

1. Sprachanalysesystem mit Eingabe über eine Mikrofonanordnung, umfassend:

eine Spracheingabeeinheit (1) zur Eingabe von Sprache in einer Vielzahl von Kanälen unter Verwendung einer Mikrofonanordnung, welche durch eine Vielzahl von Mikrofonen gebildet wird; und

eine Frequenzanalyseeinheit (2) zur Analyse der Spracheingabe jedes Kanals, welche durch die Spracheingabeeinheit eingegeben wird, und zum Erhalten von Bandpass-Signalformen jedes Kanals, wobei jede Bandpass-Signalform eine Signalform für jede Frequenzbandbreite ist;

gekennzeichnet durch
eine Schallquellenpositions-Sucheinheit (3) zur Berechnung einer Bandpass-Leistungsverteilung als Funktion der Schallquellenposition oder -richtung für jede Frequenzbandbreite aus den Bandpass-Signalformen für jede **durch** die Frequenzanalyseeinheit erhaltene Frequenzbandbreite, zur Synthetisierung von berechneten Bandpass-Leistungsverteilungen für eine Vielzahl von Frequenzbandbreiten, und zur Abschätzung einer Schallquellenposition oder -richtung aus einer synthetisierten Bandpass-Leistungsverteilung.

2. System nach Anspruch 1, ferner umfassend:

eine Sprachparameter-Extraktionseinheit (4) zur Extrahierung eines Sprachparameters aus der Bandpass-Leistungsverteilung für jede Frequenzbandbreite, die durch die Schallquellenpositions-Sucheinheit berechnet wurde, gemäß der Schallquellenposition oder -richtung, die durch die Schallquellenpositions-Sucheinheit abgeschätzt wurde.

3. System nach Anspruch 2, ferner umfassend:

eine Spracherkennungseinheit (5) zur Gewinnung eines Spracherkennungsresultats durch Vergleichen des durch die Sprachparameter-Extraktionseinheit extrahierten Sprachparameters mit einem Erkennungslexikon.

4. System nach einem der Ansprüche 1, 2 und 3, wobei die Schallquellenpositions-Sucheinheit umfasst:

eine Schallquellenpositions-Abschätzungseinheit niedriger Auflösung zur Abschätzung einer groben Schallquellenposition oder -richtung, durch Minimierung der Ausgangsleistung der Mikrofonanordnung unter Randbedingungen, dass Antworten der Mikrofonanordnung für eine Vielzahl von Richtungen oder Positionen konstant gehalten werden; und

eine Schallquellenpositions-Abschätzungseinheit hoher Auflösung zur Abschätzung einer genauen Schallquellenposition oder -richtung in der Nähe der groben Schallquellenposition- oder richtung, welche durch die Schallquellenpositions-Abschätzungseinheit niedriger Auflösung abgeschätzt wurde, durch Minimierung der Ausgangsleistung der Mikrofonanordnung unter Randbedingungen, dass eine Antwort der Mikrofonanordnung für eine Richtung oder Position konstant gehalten werden soll, wobei die Sprachparameter-Extraktionseinheit den Sprachparameter für Spracherkennung entsprechend der genauen Schallquellenposition oder -richtung extrahiert.

5. System nach Anspruch 3, wobei die Frequenzanalyseeinheit die Bandpass-Signalformen für jeden Kanal unter Verwendung einer Bandpass-Filterbank gewinnt.

6. System nach Anspruch 3, wobei die Schallquellenpositions-Sucheinheit die Bandpass-Leistungsverteilung für jede Frequenzbandbreite dadurch berechnet, dass die Bandpassleistung für jede Frequenzbandbreite berechnet wird, in jeder einer Vielzahl von angenommenen Schallquellenpositionen oder - Richtungen innerhalb eines vorgeschriebenen Suchbereichs.

7. System nach Anspruch 3, wobei die Schallquellenpositions-Sucheinheit die Bandpass-Leistungsverteilung für jede Frequenzbandbreite unter Verwendung einer Filterfunktions-Konfiguration berechnet, welche eine Vielzahl von Verzögerungsleitungs-Abgriffen für jeden Kanal hat.

8. System nach Anspruch 3, wobei die Schallquellenpositions-Sucheinheit die Bandpass-Leistungsverteilung für jede Frequenzbandbreite unter Verwendung eines Verfahrens minimaler Varianz berechnet, um die Ausgangsleistung der Mikrofonanordnung unter Randbedingungen zu minimieren, dass die Antwort der Mikrofonanordnung für eine Richtung oder Position konstant gehalten wird.

9. System nach Anspruch 3, wobei die Sprachparameter-Extraktionseinheit die Bandpass-Leistungsverteilung für jede Frequenzbandbreite, die durch die Schallquellenpositions-Sucheinheit für die von der Schallquellenpositions-Sucheinheit abgeschätzten Schallquellenposition oder -richtung berechnet wurde, direkt als Sprachparameter extrahiert.

10. System nach Anspruch 3, wobei die Schallquellenpositions-Sucheinheit die berechneten Bandpass-Leistungsverteilungen für eine Vielzahl von Frequenzbandbreiten berechnet, indem die berechneten Bandpass-Leistungsverteilungen mit jeweiligen Gewichten gewichtet werden und die gewichteten Bandpass-Leistungsverteilungen summiert werden.

11. System nach Anspruch 3, wobei die Schallquellenpositions-Sucheinheit die Schallquellenposition oder -richtung abschätzt, indem einen Spitze in der

synthetisierten Bandpass-Leistungsverteilung erfasst wird, und eine Position oder Richtung entsprechend einer erfassten Spitze als Schallquellenposition oder -richtung eingestellt wird.

12. Sprachanalyseverfahren mit Eingabe über eine Mikrofonanordnung, umfassend die Schritte:

Eingeben von Sprache in einer Vielzahl von Kanälen unter Verwendung einer Mikrofonanordnung, welche durch eine Vielzahl von Mikrofonen gebildet ist; und

Analysieren der Spracheingabe jedes Kanals, welche durch den Eingabeschritt eingegeben wurde, und Gewinnung von Bandpass-Signalformen für jeden Kanal, wobei jede Bandpass-Signalform eine Signalform für jede Frequenzbandbreite ist;

gekennzeichnet durch
Berechnen einer Bandpass-Leistungsverteilung als Funktion der Schallquellenposition oder -richtung für jede Frequenzbandbreite aus den Bandpass-Signalformen für jede **durch** den Analyseschritt gewonnene Frequenzbandbreite, Synthetisieren der berechneten Bandpass-Leistungsverteilungen für eine Vielzahl von Frequenzbandbreiten, und Abschätzen einer Schallquellenposition oder -richtung aus einer synthetisieren Bandpass-Leistungsverteilung.

13. Verfahren nach Anspruch 12, ferner umfassend den Schritt:

Extrahieren eines Sprachparameters aus der Bandpass-Leistungsverteilung für jede durch den Berechnungsschritt berechnete Frequenzbandbreite, gemäß der durch den Berechnungsschritt abgeschätzten Schallquellenposition oder -richtung.

14. Verfahren nach Anspruch 13, ferner umfassend den Schritt:

Gewinnen eines Spracherkennungsresultats durch Vergleichen des durch den Extrahierungsschritt extrahierten Sprachparameters mit einem Erkennungslexikon.

15. Verfahren nach einem der Ansprüche 12, 13, und 14, wobei der Berechnungsschritt die Schritte umfasst:

einen Schallquellenpositions-Abschätzungsschritt niedriger Auflösung zur Abschätzung einer groben Schallquellenposition oder -richtung durch Minimierung der Ausgangsleistung der Mikrofonanordnung unter Randbedingungen, dass Antworten der Mikrofonanordnung für eine Vielzahl von Richtungen oder Positionen konstant gehalten werden sollen; und

einen Schallquellenpositions-Abschätzungsschritt hoher Auflösung zur Abschätzung einer genauen Schallquellenposition oder -richtung in der Nähe der groben Schallquellenposition oder -richtung, die durch den Schallquellenpositions-Abschätzungsschritt niedriger Auflösung abgeschätzt wurde, durch Minimierung der Ausgangsleistung der Mikrofonanordnung unter Randbedingungen, dass eine Antwort der Mikrofonanordnung für eine Richtung oder Position konstant gehalten werden soll, wobei der Extraktionsschritt den Sprachparameter für die Spracherkennung gemäß der genauen Schallquellenposition oder -richtung extrahiert.

16. Verfahren nach Anspruch 14, wobei der Analyseschritt die Bandpass-Signalformen für jeden Kanal unter Verwendung einer Bandpass-Filterbank gewinnt.

17. Verfahren nach Anspruch 14, wobei der Berechnungsschritt die Bandpass-Leistungsverteilung für jede Frequenzbandbreite berechnet, durch Berechnen einer Bandpassleistung für jede Frequenzbandbreite, in jeder einer Vielzahl von angenommenen Schallquellenpositionen oder -richtungen innerhalb eines vorgeschriebenen Suchbereichs.

18. Verfahren nach Anspruch 14, wobei der Berechnungsschritt die Bandpass-Leistungsverteilung für jede Frequenzbandbreite unter Verwendung einer Filterfunktionskonfiguration berechnet, welche eine Vielzahl von Verzögerungsleitungsabgriffen für jeden Kanal hat.

19. Verfahren nach Anspruch 14, wobei der Berechnungsschritt die Bandpass-Leistungsverteilung für jede Frequenzbandbreite unter Verwendung eines Verfahrens minimaler Varianz berechnet, zur Minimierung der Ausgangslei-

stung der Mikrofonanordnung unter Randbedingungen, dass eine Antwort der Mikrofonanordnung für eine Richtung oder Position konstant gehalten werden soll.

20. Verfahren nach Anspruch 14, wobei der Extraktionsschritt die Bandpass-Leistungsverteilung für jede Frequenzbandbreite, welche durch den Berechnungsschritt für die durch den Berechnungsschritt abgeschätzte Schallquellenposition oder -Richtung berechnet wurde, direkt als Sprachparameter extrahiert.

21. Verfahren nach Anspruch 14, wobei der Berechnungsschritt die berechneten Bandpass-Leistungsverteilungen für eine Vielzahl von Frequenzbandbreiten synthetisiert, durch Gewichtung der berechneten Bandpass-Leistungsverteilungen mit entsprechenden Gewichten, und der Summierung der gewichteten Bandpass-Leistungsverteilungen.

22. Verfahren nach Anspruch 14, wobei der Berechnungsschritt die Schallquellenposition oder -richtung durch Erfassen einer Spitze in der synthetisierten Bandpass-Leistungsverteilung und Einstellen einer Position oder Richtung entsprechend einer erfassten Spitze als Schallquellenposition oder -richtung abschätzt.

Revendications

1. Système d'analyse de la parole du type à entrée par réseau de microphones, comprenant :

une unité d'entrée de parole (1) pour entrer des paroles dans une pluralité de canaux en utilisant un réseau de microphones qui est formé par une pluralité de microphones ; et
une unité d'analyse de fréquence (2) pour analyser une parole d'entrée de chaque canal qui est entrée au moyen de l'unité d'entrée de parole et pour obtenir des formes d'onde passe-bande pour chaque canal, chaque forme d'onde passe-bande étant une forme d'onde pour chaque largeur de bande de fréquences,

caractérisé par :

une unité de recherche de position de source de son (3) pour calculer une distribution de puissance passe-bande en tant que fonction d'une position ou d'une direction de source de son pour chaque largeur de bande de fréquences à partir des formes d'ondes passe-bande pour chaque largeur de bande de fréquences comme obtenu au moyen de l'unité d'analyse de fréquence, pour synthétiser des distributions de puissance passe-bande calculées pour une pluralité de largeurs de bande de fréquences et pour estimer une position ou une direction de source de son à partir d'une distribution de puissance passe-bande synthétisée.

2. Système selon la revendication 1, comprenant en outre :

une unité d'extraction de paramètre de parole (4) pour extraire un paramètre de parole à partir de la distribution de puissance passe-bande pour chaque largeur de bande de fréquences calculée au moyen de l'unité de recherche de position de source de son conformément à la position ou à la direction de source de son estimée au moyen de l'unité de recherche de position de source de son.

3. Système selon la revendication 2, comprenant en outre :

une unité de reconnaissance de la parole (5) pour obtenir un résultat de reconnaissance de la parole en faisant correspondre le paramètre de parole qui est extrait au moyen de l'unité d'extraction de paramètre de parole avec un dictionnaire de reconnaissance.

4. Système selon l'une quelconque des revendications 1, 2 et 3, dans lequel l'unité de recherche de position de source de son inclut :

une unité d'estimation de position de source de son basse résolution pour estimer une position ou une direction de source de son grossière, en minimisant une puissance de sortie du réseau de microphones sous des contraintes consistant en ce que des réponses du réseau de microphones pour une pluralité de directions ou de positions doivent être maintenues constantes ; et
une unité d'estimation de position de source de son haute résolution pour estimer une position ou une direction de source de son précise au voisinage de la position ou de la direction de source de son grossière estimée au moyen de l'unité d'estimation de position de source de son basse résolution, en minimisant la puissance

de sortie du réseau de microphones sous des contraintes consistant en ce qu'une réponse du réseau de microphones pour une direction ou une position doit être maintenue constante, où l'unité d'extraction de paramètre de parole extrait le paramètre de parole pour une reconnaissance de parole en fonction de la position ou de la direction de source de son précise.

5. Système selon la revendication 3, dans lequel l'unité d'analyse de fréquence obtient les formes d'onde passe-bande pour chaque canal en utilisant un groupe de filtres passe-bande.

6. Système selon la revendication 3, dans lequel l'unité de recherche de position de source de son calcule la distribution de puissance passe-bande pour chaque largeur de bande de fréquences, en calculant une puissance passe-bande pour chaque largeur de bande de fréquences, pour chacune d'une pluralité de positions ou de directions de source de son prises à l'intérieur d'une plage de recherche prescrite.

7. Système selon la revendication 3, dans lequel l'unité de recherche de position de source de son calcule la distribution de puissance passe-bande pour chaque largeur de bande de fréquences en utilisant une configuration de fonction de filtre comportant une pluralité de connexions intermédiaires de ligne de retard pour chaque canal.

8. Système selon la revendication 3, dans lequel l'unité de recherche de position de source de son calcule la distribution de puissance passe-bande pour chaque largeur de bande de fréquences en utilisant un procédé par variance minimum pour minimiser une puissance de sortie du réseau de microphones sous des contraintes consistant en ce qu'une réponse du réseau de microphones pour une direction ou une position doit être maintenue constante.

9. Système selon la revendication 3, dans lequel l'unité d'extraction de paramètre de parole extrait la distribution de puissance passe-bande pour chaque largeur de bande de fréquences calculée au moyen de l'unité de recherche de position de source de son pour la position ou la direction de source de son estimée au moyen de l'unité de recherche de position de source de son directement en tant que paramètre de parole.

10. Système selon la revendication 3, dans lequel l'unité de recherche de position de source de son synthétise les distributions de puissance passe-bande calculées pour une pluralité de largeurs de bande de fréquences en pondérant les distributions de puissance passe-bande calculées à l'aide de poids respectifs et en sommant les distributions de puissance passe-bande pondérées.

11. Système selon la revendication 3, dans lequel l'unité de recherche de position de source de son estime la position ou la direction de source de son en détectant une crête dans la distribution de puissance passe-bande synthétisée et en établissant une position ou une direction qui correspond à une crête détectée en tant que position ou direction de source de son.

12. Procédé d'analyse de la parole du type à entrée par réseau de microphones, comprenant les étapes de :

entrée de paroles dans une pluralité de canaux en utilisant un réseau de microphones qui est formé par une pluralité de microphones ; et
analyse d'une parole d'entrée de chaque canal qui est entrée au moyen de l'étape d'entrée et obtention d'une forme d'onde passe-bande pour chaque canal, chaque forme d'onde passe-bande étant une forme d'onde pour chaque largeur de bande de fréquences,

caractérisé par :
le calcul d'une distribution de puissance passe-bande en tant que fonction d'une position ou d'une direction de source de son pour chaque largeur de bande de fréquences à partir des formes d'onde passe-bande pour chaque largeur de bande de fréquences comme obtenu au moyen de l'étape d'analyse, la synthèse de distributions de puissance passe-bande calculées pour une pluralité de largeurs de bande de fréquences et l'estimation d'une position ou d'une direction de source de son à partir d'une distribution de puissance passe-bande synthétisée.

13. Procédé selon la revendication 12, comprenant en outre l'étape de :

extraction d'un paramètre de parole à partir de la distribution de puissance passe-bande pour chaque largeur de bande de fréquences calculée au moyen de l'étape de calcul, conformément à la position ou à la direction de source de son qui est estimée au moyen de l'étape de calcul.

14. Procédé selon la revendication 13, comprenant en outre l'étape de :

obtention d'un résultat de reconnaissance de parole en faisant correspondre le paramètre de parole qui est extrait au moyen de l'étape d'extraction avec un dictionnaire de reconnaissance.

15. Procédé selon l'une quelconque des revendications 12, 13 et 14, dans lequel l'étape de calcul inclut les étapes de :

une étape d'estimation de position de source de son basse résolution pour estimer une position ou une direction de source de son grossière en minimisant une puissance de sortie du réseau de microphones sous des contraintes consistant en ce que des réponses du réseau de microphones pour une pluralité de directions ou de positions doivent être maintenues constantes ; et
une étape d'estimation de position de source de son haute résolution pour estimer une position ou une direction de source de son précise au voisinage d'une position ou d'une direction de source de son grossière estimée au moyen de l'étape d'estimation de position de source de son basse résolution, en minimisant la puissance de sortie du réseau de microphones sous des contraintes consistant en ce qu'une réponse du réseau de microphones pour une direction ou une position doit être maintenue constante, où l'étape d'extraction extrait le paramètre de parole pour une reconnaisse de parole conformément à la position ou à la direction de source de son précise.

16. Procédé selon la revendication 14, dans lequel l'étape d'analyse obtient les formes d'onde passe-bande pour chaque canal en utilisant un groupe de filtres passe-bande.

17. Procédé selon la revendication 14, dans lequel l'étape de calcul calcule la distribution de puissance passe-bande pour chaque largeur de bande de fréquences en calculant une puissance passe-bande pour chaque largeur de bande de fréquences pour chacune d'une pluralité de positions ou de directions de source de son prises à l'intérieur d'une plage de recherche prescrite.

18. Procédé selon la revendication 14, dans lequel l'étape de calcul calcule la distribution de puissance passe-bande pour chaque largeur de bande de fréquences en utilisant une configuration de fonction de filtre comportant une pluralité de connexions intermédiaires de ligne de retard pour chaque canal.

19. Procédé selon la revendication 14, dans lequel l'étape de calcul calcule la distribution de puissance passe-bande pour chaque largeur de bande de fréquences en utilisant un procédé par variance minimum pour minimiser une puissance de sortie du réseau de microphones sous des contraintes consistant en ce qu'une réponse du réseau de microphones pour une direction et une position doit être maintenue constante.

20. Procédé selon la revendication 14, dans lequel l'étape d'extraction extrait la distribution de puissance passe-bande pour chaque largeur de bande de fréquences calculée au moyen de l'étape de calcul pour la position ou la direction de source de son estimée au moyen de l'étape de calcul directement en tant que paramètre de parole.

21. Procédé selon la revendication 14, dans lequel l'étape de calcul synthétise les distributions de puissance passe-bande calculées pour une pluralité de largeurs de bande de fréquences en pondérant les distributions de puissance passe-bande calculées avec des poids respectifs et en sommant les distributions de puissance passe-bande pondérées.

22. Procédé selon la revendication 14, dans lequel l'étape de calcul estime la position ou la direction de source de son en détectant une crête dans la distribution de puissance passe-bande synthétisée et en établissant une position ou une direction qui correspond à une crête détectée en tant que position ou direction de source de son.

FIG.1
PRIOR ART

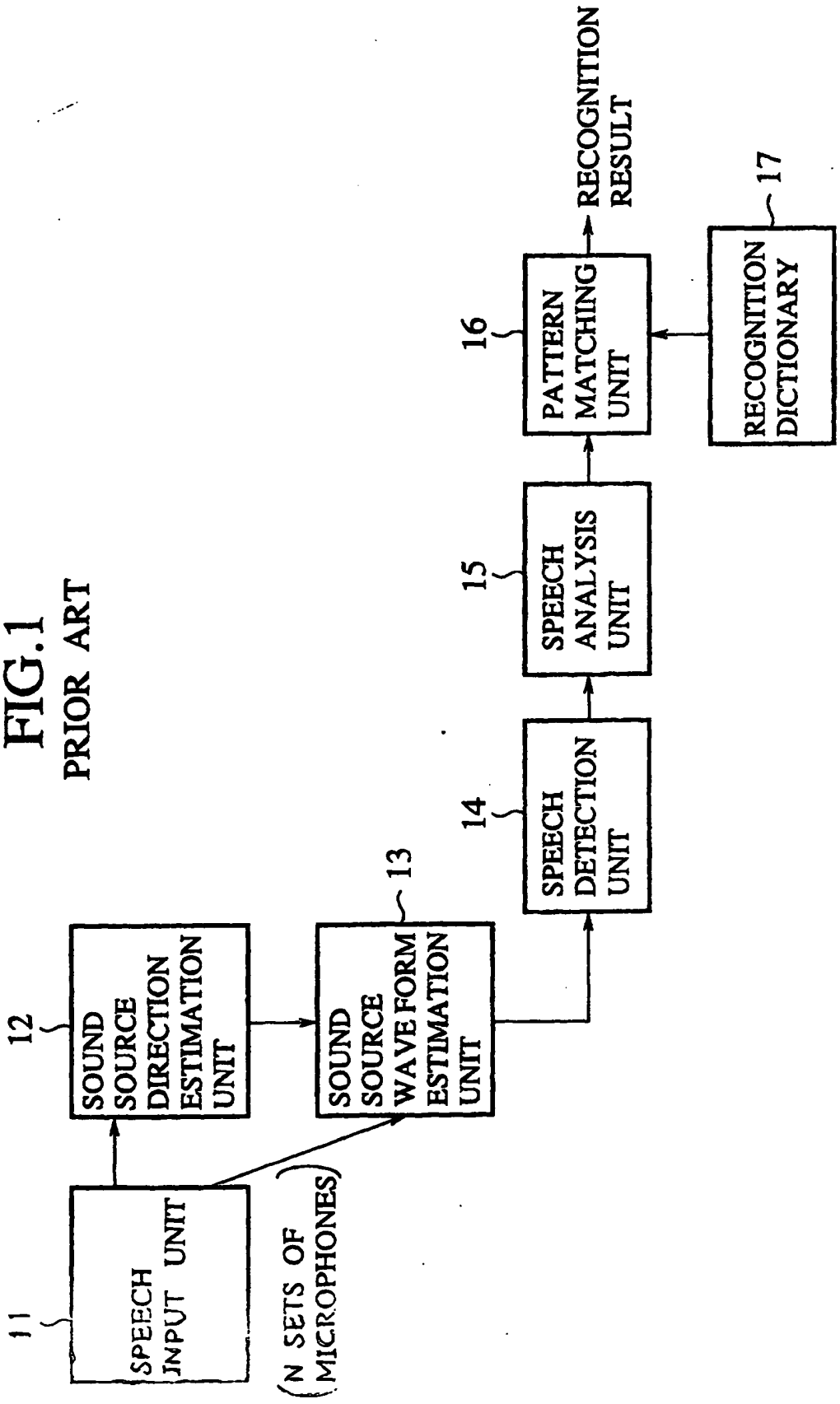


FIG.2
PRIOR ART

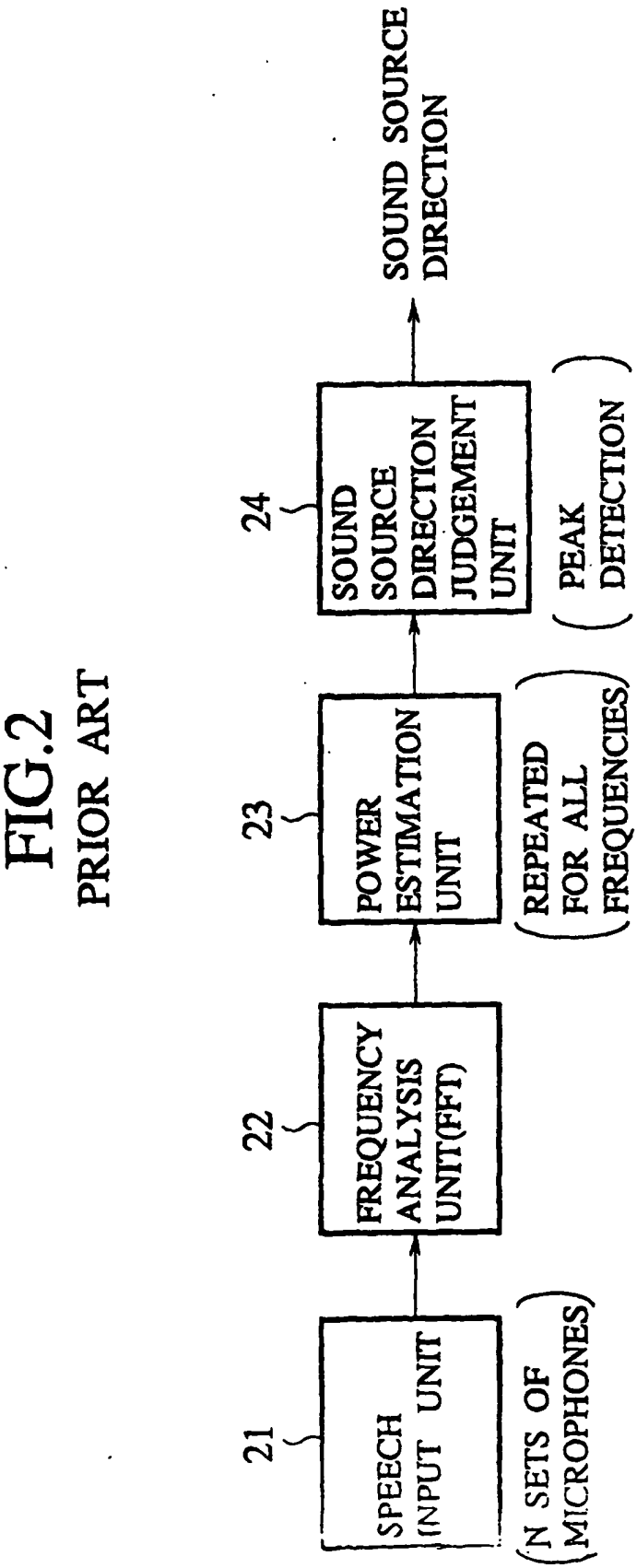


FIG.3

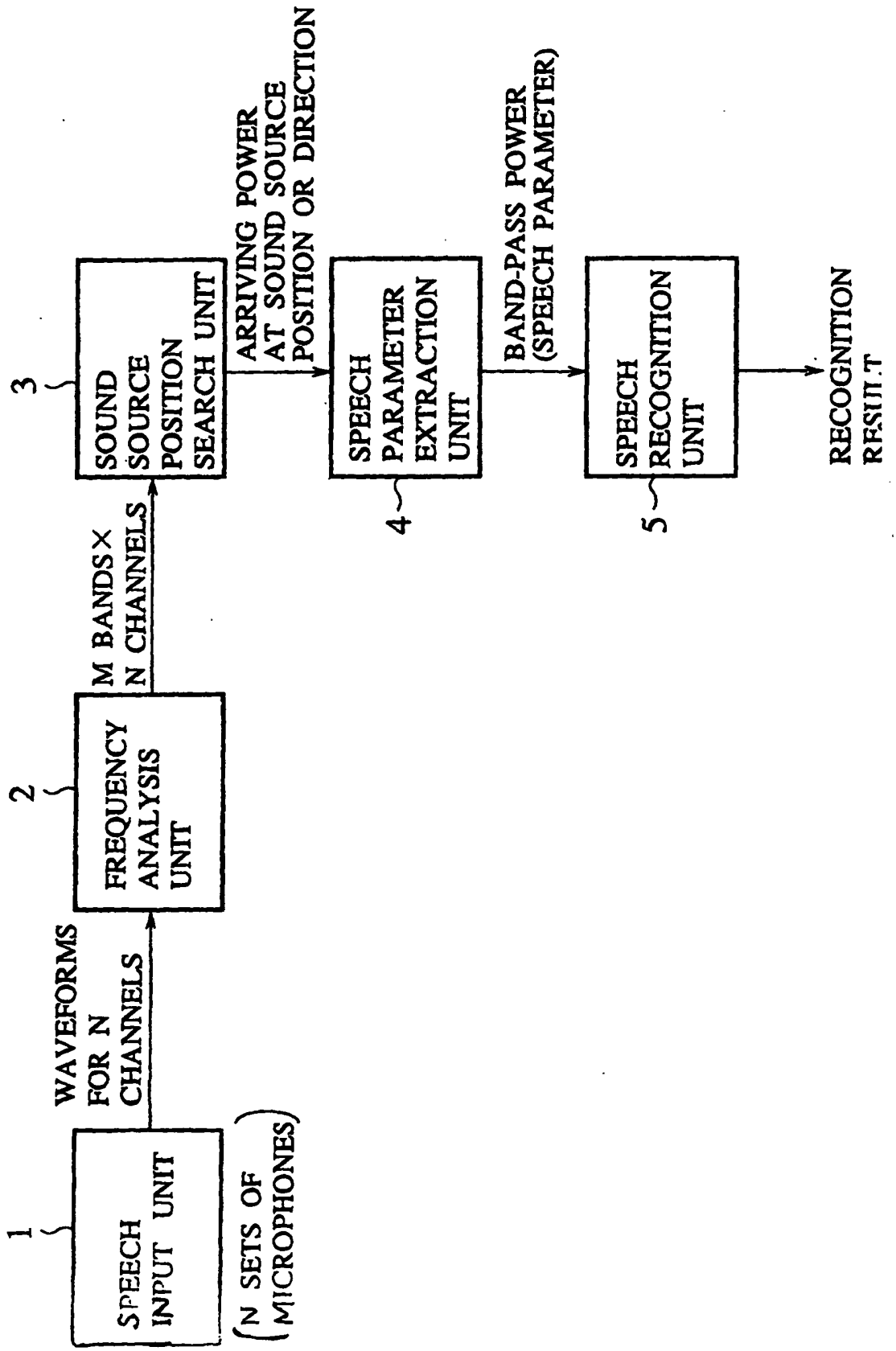

1
SPEECH INPUT UNIT
(N SETS OF MICROPHONES)
WAVEFORMS FOR N CHANNELS
2
FREQUENCY ANALYSIS UNIT
M BANDS× N CHANNELS
3
SOUND SOURCE POSITION SEARCH UNIT
ARRIVING POWER AT SOUND SOURCE POSITION OR DIRECTION
4
SPEECH PARAMETER EXTRACTION UNIT
BAND-PASS POWER (SPEECH PARAMETER)
5
SPEECH RECOGNITION UNIT
RECOGNITION RESULT

FIG.4

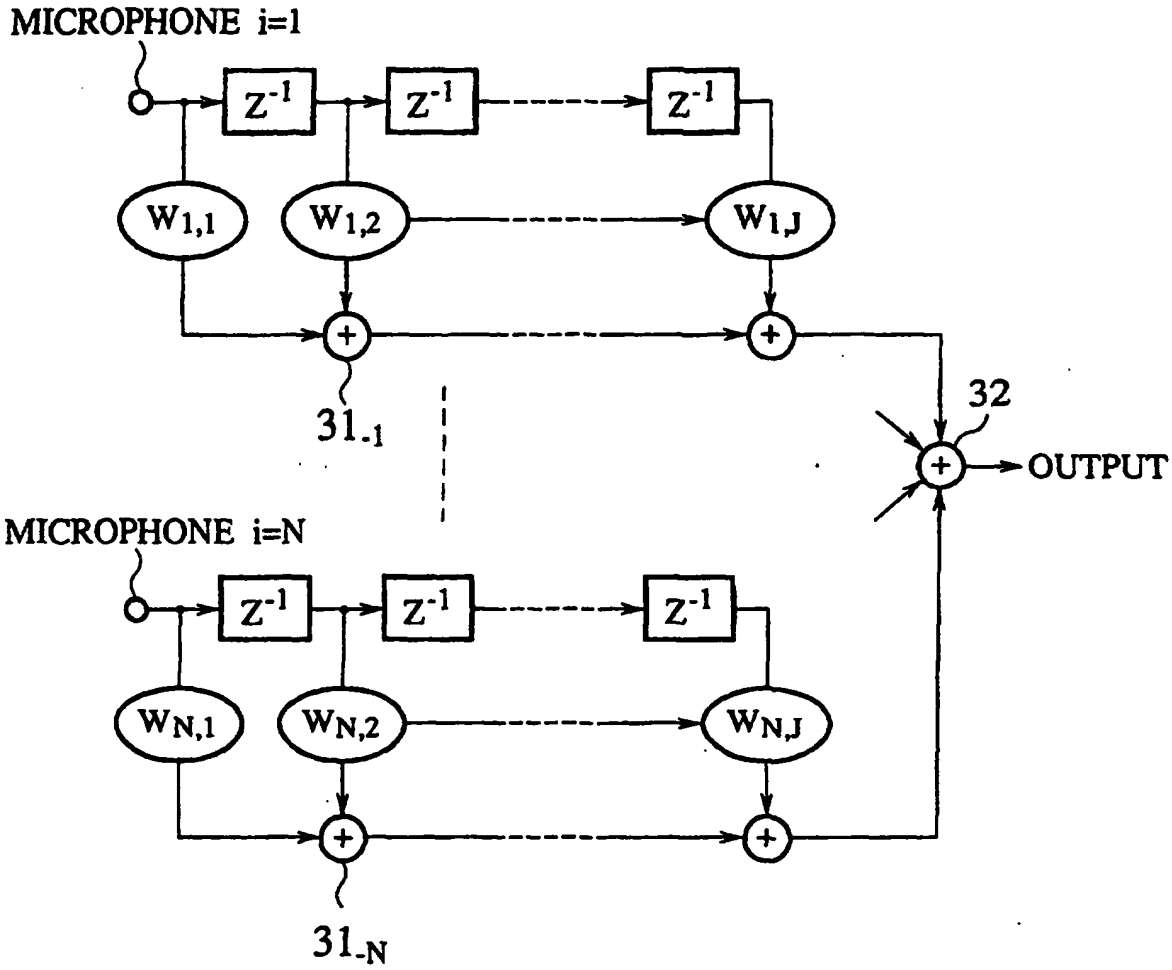

MICROPHONE i=1
Z-1
Z-1
Z-1
W1,1
W1,2
W1,J
31-1
32
OUTPUT
MICROPHONE i=N
Z-1
Z-1
Z-1
WN,1
WN,2
WN,J
31-N

FIG.5A

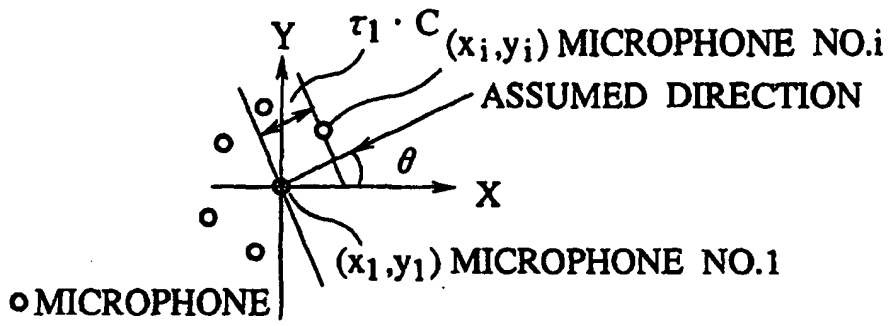


FIG.5B

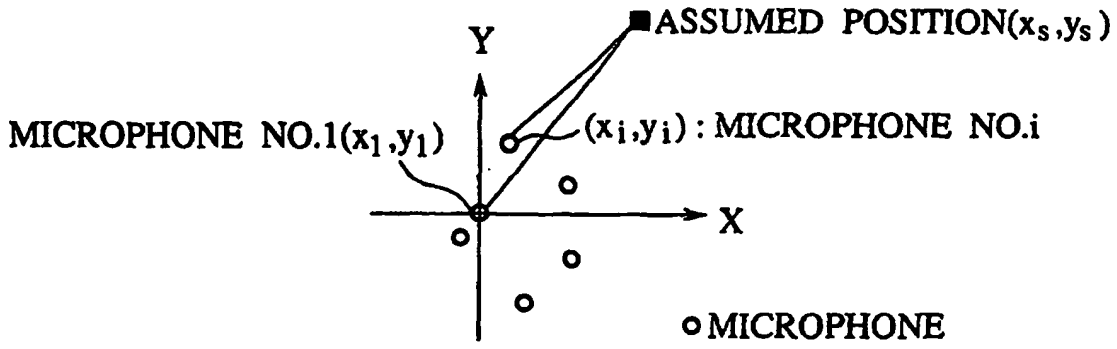


FIG.6

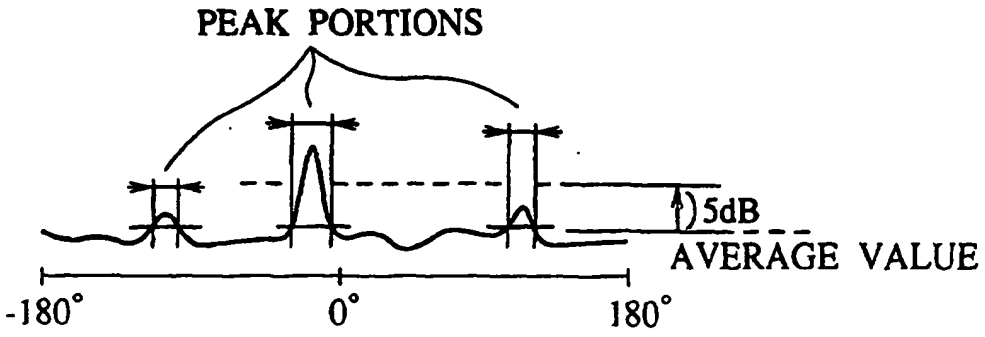


FIG.7

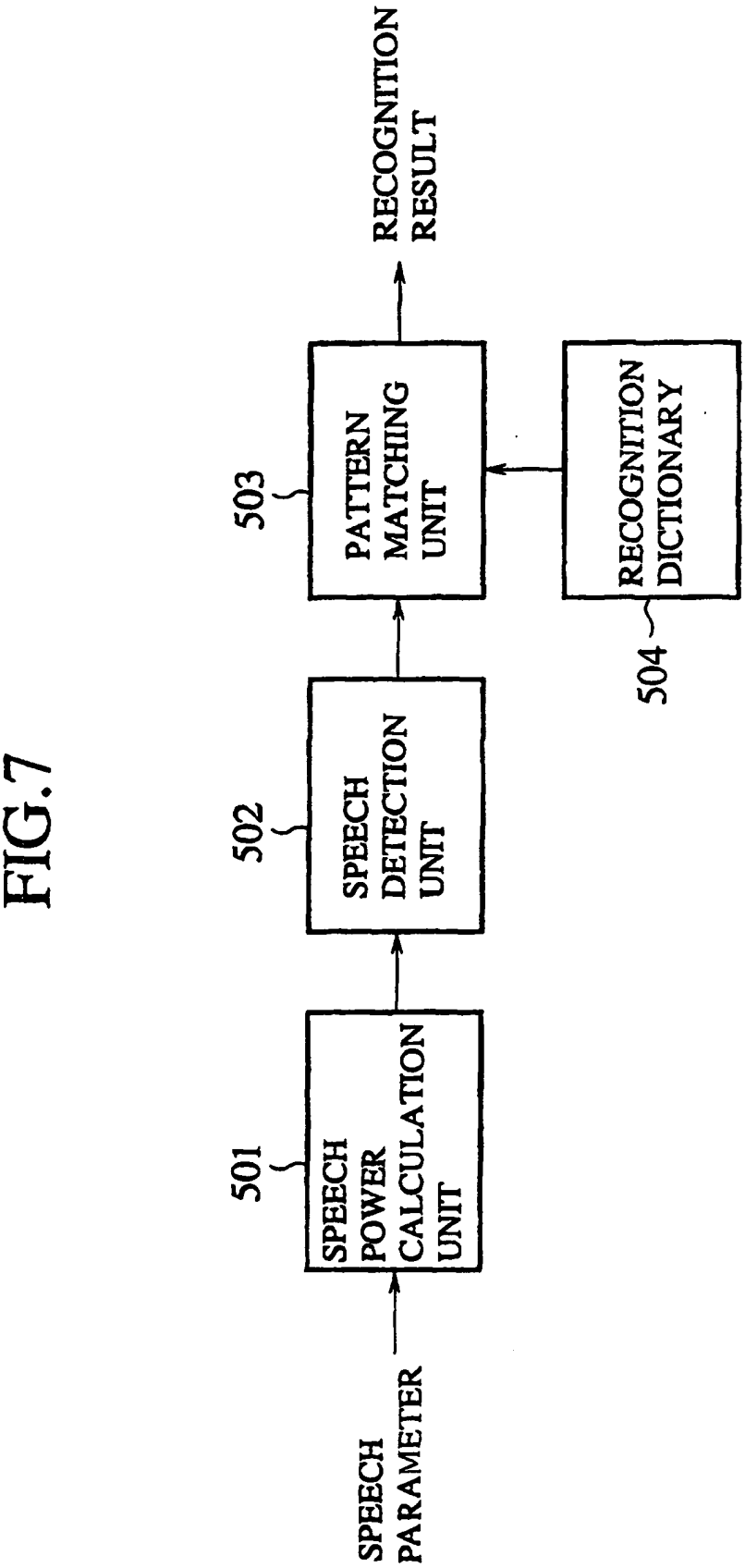

SPEECH PARAMETER
501
SPEECH POWER CALCULATION UNIT
502
SPEECH DETECTION UNIT
503
PATTERN MATCHING UNIT
504
RECOGNITION DICTIONARY
RECOGNITION RESULT

FIG.8

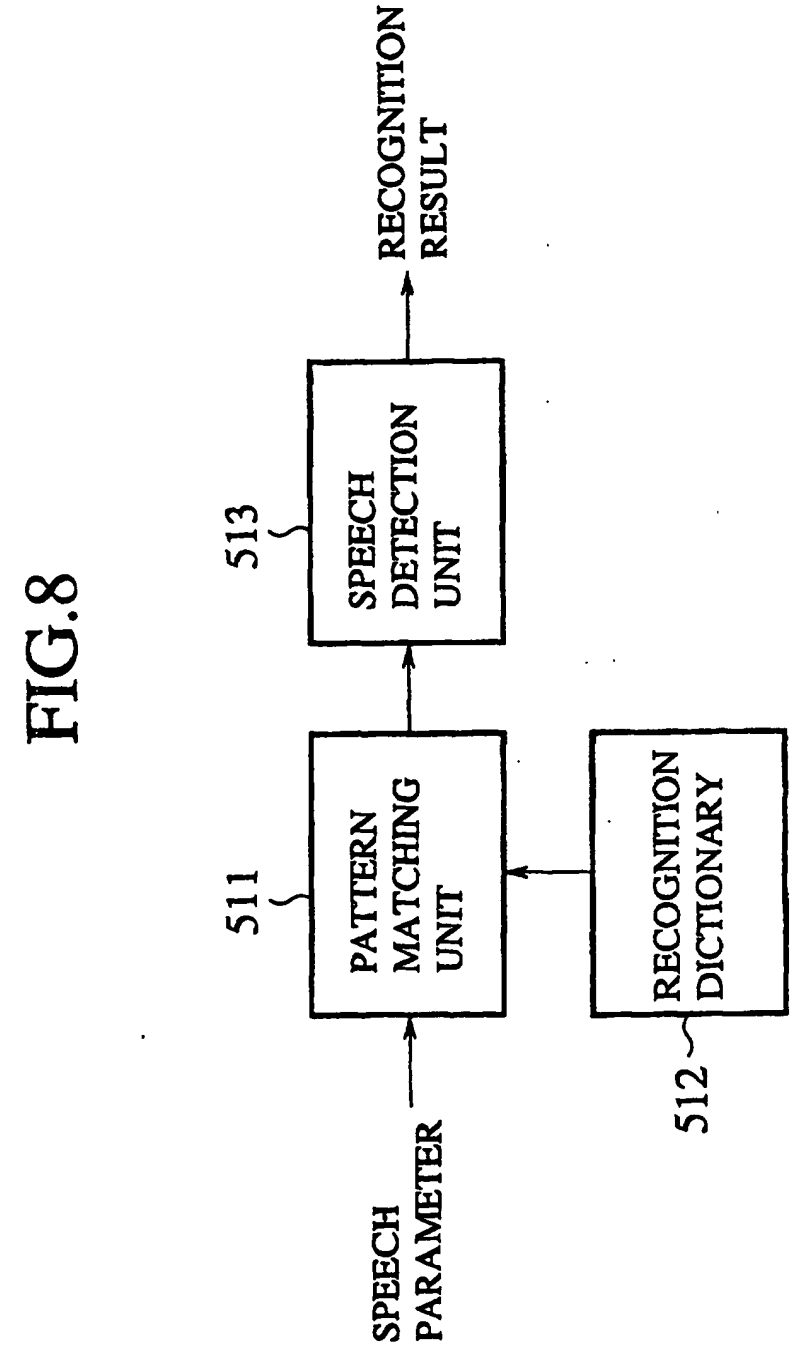

SPEECH PARAMETER
511
PATTERN MATCHING UNIT
512
RECOGNITION DICTIONARY
513
SPEECH DETECTION UNIT
RECOGNITION RESULT

FIG.9

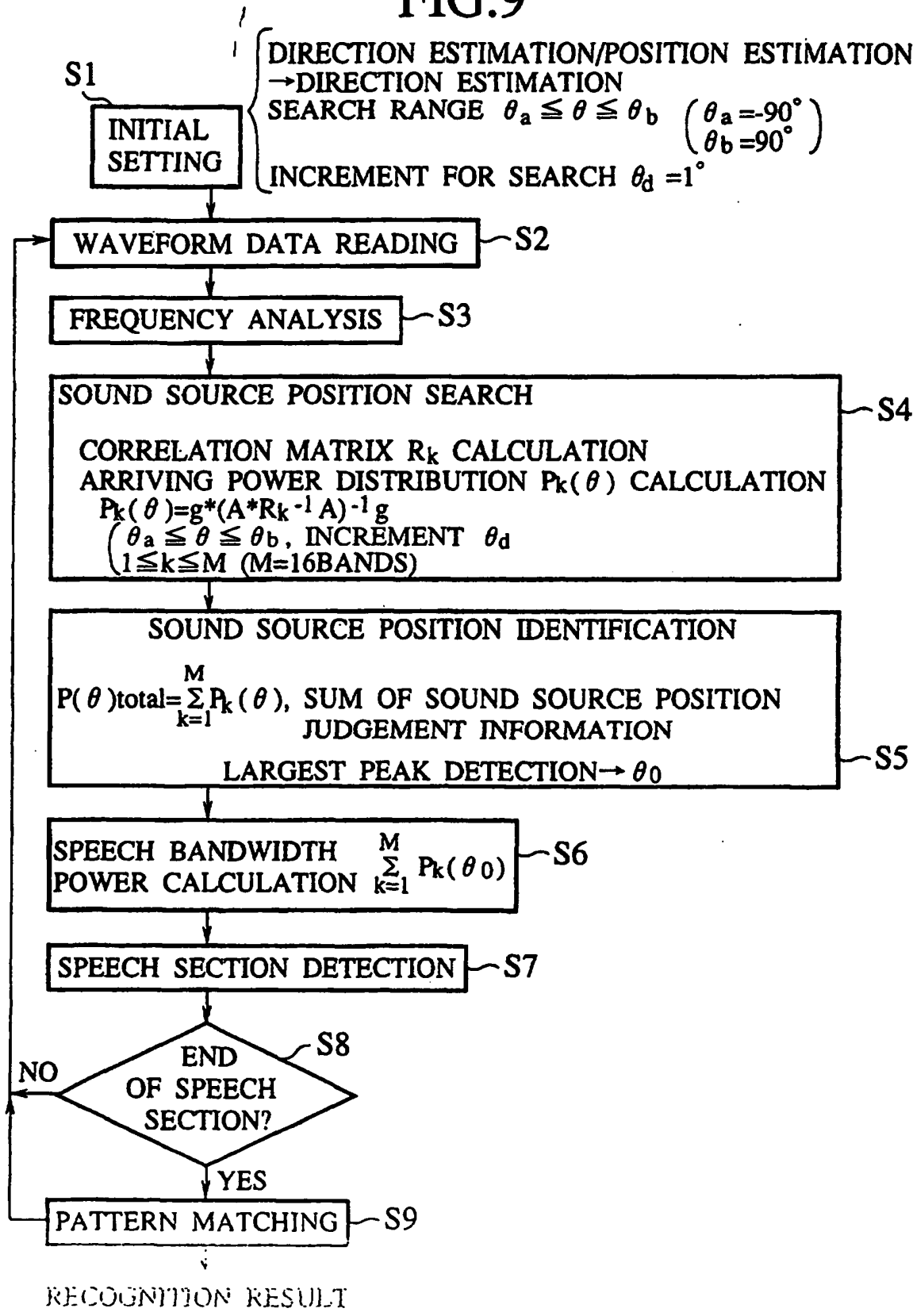

S1
INITIAL SETTING
DIRECTION ESTIMATION/POSITION ESTIMATION
→DIRECTION ESTIMATION
SEARCH RANGE $\theta_a \leqq \theta \leqq \theta_b$ (θ_a=-90°, θ_b=90°)
INCREMENT FOR SEARCH θ_d =1°
WAVEFORM DATA READING
S2
FREQUENCY ANALYSIS
S3
SOUND SOURCE POSITION SEARCH
S4
CORRELATION MATRIX R_k CALCULATION
ARRIVING POWER DISTRIBUTION $P_k(\theta)$ CALCULATION
$P_k(\theta)=g^*(A^*R_k^{-1}A)^{-1}g$
($\theta_a \leqq \theta \leqq \theta_b$, INCREMENT θ_d
$1 \leqq k \leqq M$ (M=16BANDS)
SOUND SOURCE POSITION IDENTIFICATION
$P(\theta)total=\sum_{k=1}^{M} P_k(\theta)$, SUM OF SOUND SOURCE POSITION JUDGEMENT INFORMATION
LARGEST PEAK DETECTION→ θ_0
S5
SPEECH BANDWIDTH POWER CALCULATION $\sum_{k=1}^{M} P_k(\theta_0)$
S6
SPEECH SECTION DETECTION
S7
END OF SPEECH SECTION?
S8
NO
YES
PATTERN MATCHING
S9
RECOGNITION RESULT

FIG.10

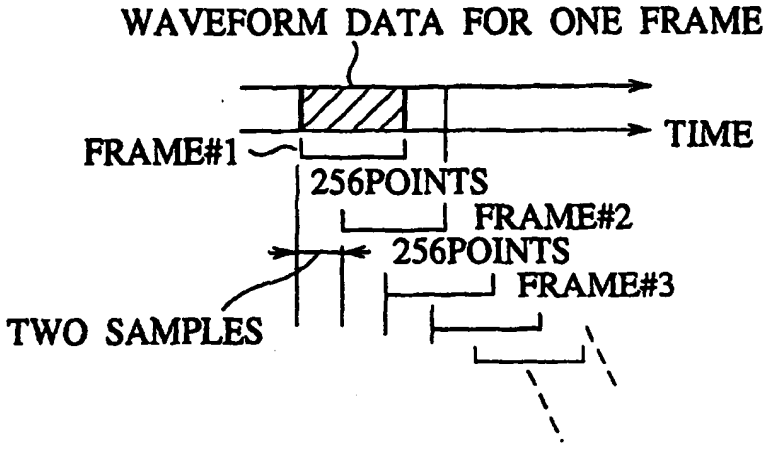


FIG.11A

HIGH RESOLUTION, COARSE SEARCH

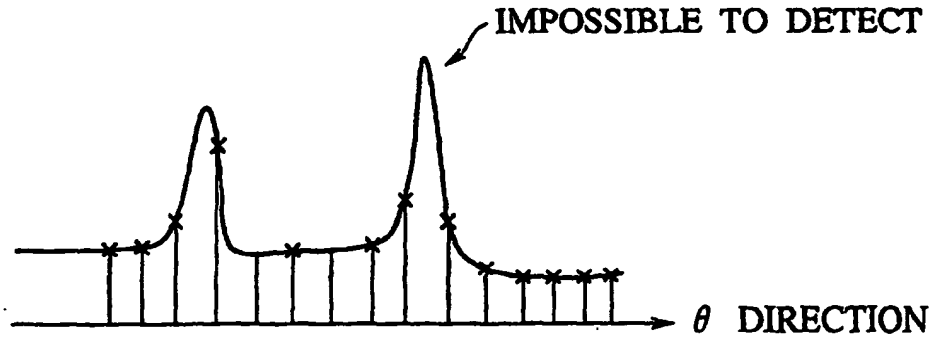


FIG.11B

LOW RESOLUTION, COARSE SEARCH

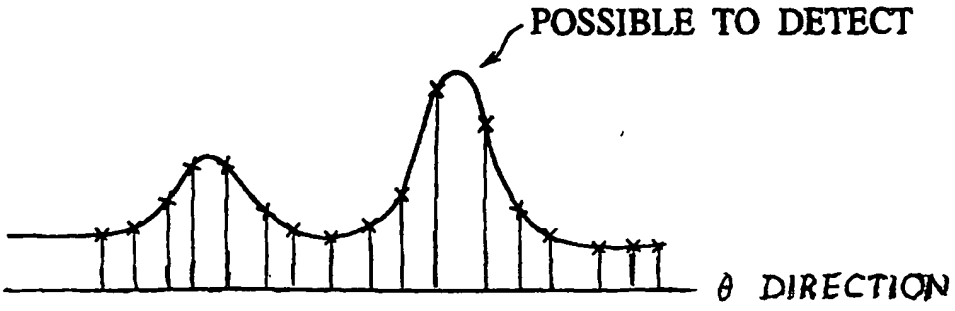


FIG.12

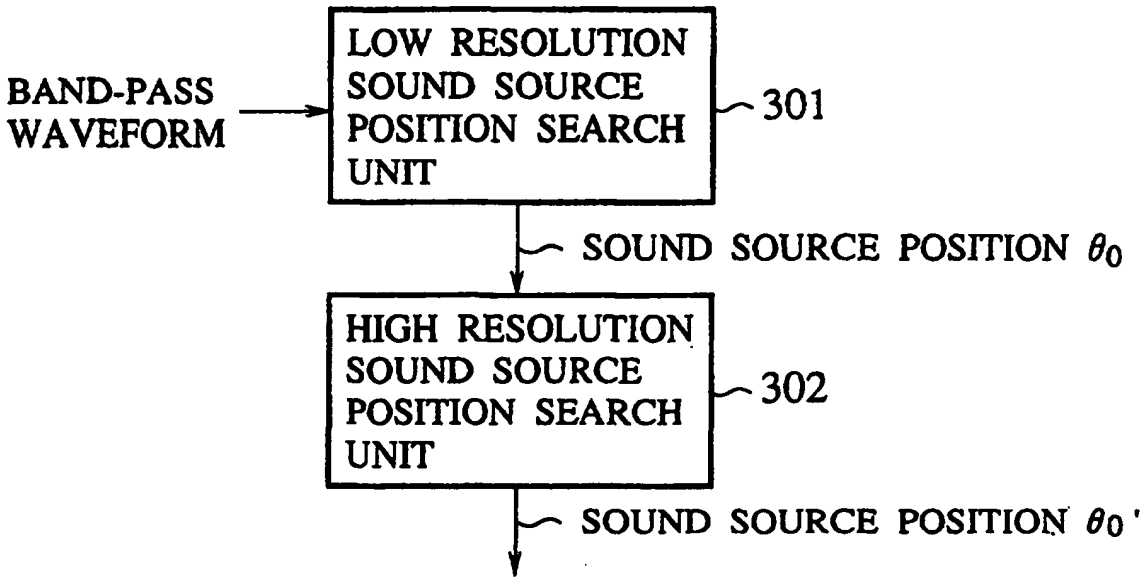
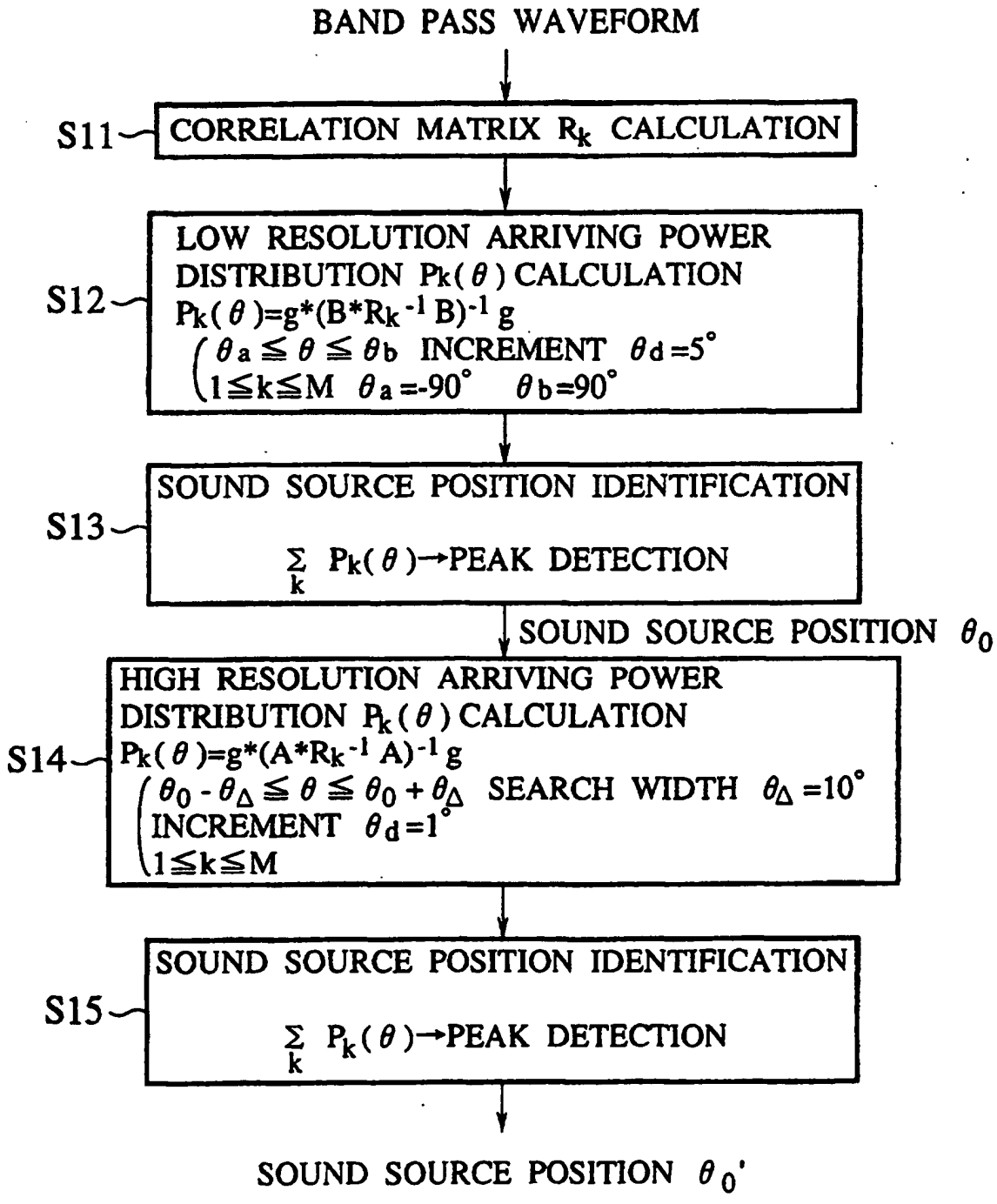

BAND-PASS WAVEFORM
LOW RESOLUTION SOUND SOURCE POSITION SEARCH UNIT
301
SOUND SOURCE POSITION θ_0
HIGH RESOLUTION SOUND SOURCE POSITION SEARCH UNIT
302
SOUND SOURCE POSITION θ_0'

FIG.13


BAND PASS WAVEFORM
S11
CORRELATION MATRIX R_k CALCULATION
S12
LOW RESOLUTION ARRIVING POWER DISTRIBUTION $P_k(\theta)$ CALCULATION
$P_k(\theta)=g^*(B^*R_k^{-1}B)^{-1}g$
$\theta_a \leqq \theta \leqq \theta_b$ INCREMENT $\theta_d=5°$
$1 \leqq k \leqq M$ $\theta_a=-90°$ $\theta_b=90°$
S13
SOUND SOURCE POSITION IDENTIFICATION
$\sum_k P_k(\theta) \rightarrow$PEAK DETECTION
SOUND SOURCE POSITION θ_0
S14
HIGH RESOLUTION ARRIVING POWER DISTRIBUTION $P_k(\theta)$ CALCULATION
$P_k(\theta)=g^*(A^*R_k^{-1}A)^{-1}g$
$\theta_0-\theta_\Delta \leqq \theta \leqq \theta_0+\theta_\Delta$ SEARCH WIDTH $\theta_\Delta=10°$
INCREMENT $\theta_d=1°$
$1 \leqq k \leqq M$
S15
SOUND SOURCE POSITION IDENTIFICATION
$\sum_k P_k(\theta) \rightarrow$PEAK DETECTION
SOUND SOURCE POSITION θ_0'